

nature

30 August 1979

Dr Boyson's provocative challenge

No-one likes cuts in higher education — but if they must come, let them be selective

There was an uneasy feeling around in the British higher education community at the time of the first Conservative budget (in June) that such belt-tightening as was called for (it included the unwelcome increase in VAT) was not the end of the matter. The Government was taking its time brooding about education and science, and was not really expected to show its hand before the Autumn. Now at least one aspect of the forthcoming white paper on education has been made clear, namely that there will be a cut of up to 6% in student intake into universities in 1980. The University Grants Committee, on the recommendation of the Department of Education and Science, has alerted universities not to make 1980 commitments beyond 94% of their 1979 levels.

Dr Rhodes Boyson, the UK Minister in charge of higher and further education, has defended the proposed cuts on the grounds that Britain now has "the most expensive higher education system in the world". This seems hard to justify in the face of recently published figures which in fact show that Britain's spending on education is the third-lowest sum per capita in the Common Market (*Euroforum*; 13 July 79). In 1975 an average of £144 was spent on education per head of population compared with figures of £277 in Denmark, £239 in Holland and £209 in Belgium. Only Ireland and Italy came lower on the list.

It is clear that Dr Rhodes Boyson, minister in charge of higher and further education, is flexing his muscles to do what politicians have traditionally shrunk away from — interfere in the sacred area of a university's autonomy to teach what subjects it wishes. He is reported to have said that he would defend keeping alive all subjects currently being taught, but "in some universities, not all". It is clear that Dr Boyson would seek to impose the least restraint on mathematicians and "hard scientists" whom he regards as in short supply, but he adds ominously that "there is no shortage in certain arts subjects", and he has apparently asked officials at the Department of Education and Science to report on subject areas which have large graduate unemployment. This, of course, makes the assumption that the present-day predilections of employers should control the future of British universities.

However it is important to see these cuts in their detailed perspective — and the number of potential university students change substantially year by year. Recently the

percentage of eighteen-year-olds who have been both able and willing to pursue a university course has remained fairly static. On the other hand, the baby boom of the late 1950s and early 1960s has raised the numbers of those in the eighteen-year-old bracket by a few percent each year. These numbers will continue to grow into the early 1980s, but thereafter will show a progressive decline until by the mid-1990s there will have been a drop of 25% on the peak value in the early 1980s.

It has been a matter of discussion for some time whether universities should be restricted in their intake for the next few years in order to 'tunnel' through the peak, whether they should be allowed to expand up to the early 1980s but then be forced to contract with declining numbers seeking admission, or whether they should be allowed to expand but then encouraged to maintain peak numbers by more vigorous recruitment programmes or by extending opportunities for continuing education, or education at a later stage in life. The last Labour government seemed to be moving towards the last option, whereas the present Conservative government looks as if it is intent on 'tunnelling'. By so doing they will of course begin to close out options on continuing education.

One further piece of perspective must be added. Following the Robbins Report, there was an enormous expansion of university places in the 1960s to meet the Robbins principle that all students who wished to go to university and who were qualified should be allowed to do so. New universities sprung up and hired staff in vast numbers, thus giving the priorities of the mid-1960s enormous emphasis in the university system. There have been continuing discreet mutterings among academics about the quality of some of the departments formed and about the low standards of entry that apply in some places. So although the university community as a whole will throw up its hands in horror at Dr Boyson trespassing on their territory and daring to declare certain preferences, some thoughtful people will be quietly curious to see whether a politician will be able to pull off what has needed to be done for some time but which academics could never bring themselves to do — to recognise that the quality is not uniform throughout the British university system and to make an effort to ensure that if there have to be cuts, they do not fall uniformly everywhere



UNCSTD: Hunger and poverty

The co-existence of the enormous potential of science and technology with the continuing poverty of the developing countries was an outrageous contradiction" which the conscience of the world should no longer tolerate, Mr Bradford Morse, administrator of the United Nations Development Programme told UNCSTD delegates last week.

"Science and technology can, indeed must be used to build a better world for posterity. Yet at the present time in almost every part of this small planet in which all of our futures are invested there is hunger, malnutrition, disease, illiteracy and despair" he said.

With UNDP frequently mentioned as one of the UN agencies expected to be made responsible for carrying out in practice many of the conference's eventual recommendations and the agency itself having indicated its willingness to do this Mr Morse's words were listened to with particular interest.

"The choice and utilisation of science and technology have fundamental implications for the pattern and direction of social and economic development in each developing country, and thus for the international distribution of economic activities" he said.

"It is therefore in the interests of the world community, while respecting the various needs and objectives of all nations, to support the efforts of developing countries to strengthen their human resources, institutions and technological capabilities so as to promote broadly-based, increased economic activities. This will inevitably strengthen the foundations for self-reliant economic and social progress, and lead gradually but surely to a more equitable basis for the progress of the world economy as a whole."

Mr Morse said that in June the Governing Council of UNDP had adopted a resolution asking him to take the measures needed, in cooperation with the organisations of the UN system, to enable UNDP to respond "promptly and effectively" to decisions taken at UNCSTD and approved by the general assembly, which may be directed to UNDP.

He had also been empowered to convene a special meeting of the governing council if it proved to be necessary to achieve this. "Thus UNDP is ready to act without delay in following-up the decision of this conference and of the general assembly if it should be called upon to do so", he said.

Mr Morse made no reference to the draft programme of action which had been put



forward by the Group of 77 countries. But he described the broad outlines of what he felt the conference could achieve. These included "agreement on firm financial commitments to make an immediate practical start, and on financial targets for the substantial resource flows needed at a later stage", as well as the establishment "within a definite time-frame of the arrangements necessary to convert such targets into actual internationally supported resource commitments."

David Dickson

Iran criticises Group of 77

Iran is not impressed. In pre-revolution days the Shah's government was hopeful about the outcome of UNCSTD. The hope was to convince the developed countries to transfer technology on more suitable terms, better commercial rates, relax some of the stringent patent laws and obtain agreement for training the local manpower with all projects involving the transfer of technology. Now, according to Dr Ali Barzegar, head of the Iranian delegation, Iran believes nothing can guarantee the transfer of technology on suitable terms.

What Iran, and indeed, what the Third World needs in general, is a science and technology generating independence and not dependence on the West. "Such science and technology cannot be imported, it can evolve", says Dr Barzegar.

In Iran, Islam plays a great part in the policy, priorities and infrastructure of science and technology. How can these priorities be safeguarded with imported technology? How can that infrastructure be imported? asks Dr Barzegar.

Again, technology in Iran must enhance the cultural values of the society and allow these values to realise their full potential. How can Iran import technology that will enhance the cultural values of Iran? That technology can only be 'home grown', it must be the product of the efforts of the Iranian people.

Transfer of technology has little or no role to play in self-development. Under the Shah, Iran obtained technology from the

West at an unparalleled rate. "What did it do for the people of Iran? Turnkey schemes, foreign trained manpower, financial resources, willingness of industrial countries to transfer technology — all this is not sufficient" says Dr Barzegar.

"We had all this in Iran and nothing has happened. In fact there has been a serious debt in terms of pollution, disparity between rich and the poor in Iran, and erosion of our cultural values. So, we can say with conviction that, in the final analysis the third world stands to lose from the transfer of technology."

The belief of the Group of 77 in some type of 'Japanese miracle' is a major fallacy. Even if the Third World societies stood to gain from the transfer of technology it could only be at the expense of their cultural integrity and cultural values. "Is not the Japan of today more close to America, than to its origins?" asks Dr Barzegar.

Ziauddin Sardar

●The Arab Group is presenting a strong protest note to the UN Secretary General, Kurt Waldheim, about the timing of UNCSTD. The conference clashes with Ramadan, the Muslim month of fasting.

Mr Ali Zainul Abedin Farah, a member of the UAF delegation, said that "we like to spend Ramadan with our families, fasting and worshipping". The delegates who have had to attend UNCSTD have been deprived of the blessings of one of the holiest months of the Islamic calendar.

The protest note asks the UN to make sure that such "serious errors" are not repeated in the future.

At their second meeting, the Group also decided that all discussions carried out by the members of the Group should be in Arabic. That means that contributions to plenary sessions and committee meeting from the Arab group will be made in Arabic. The Arab league came prepared for such a resolution as its paper is available only in Arabic.

The Arab Group has made repeated declarations that they stand firmly behind the Group of 77. However, some evidence suggests that there are small factions within the group and that the support for the Group of 77 demands is not unanimous.

UNCSTD at half time: North stands firm

UNCSTD is only half way through. Yet already it seems clear that whatever emerges from the conference at the end of this week will be less a function of what the developing countries want than of what the developed countries are prepared to give them — and that this is unlikely to be very substantial.

The turning point came last Friday. Up to that point, the main focus of discussion in the working groups had been the proposal put forward by the Group of 77 that the conference should agree to establish a new financing mechanism for science and technology for development, based largely on "automatic" contributions from developed countries calculated according to a pre-agreed formula, and scheduled to reach \$2 billion by 1985 and \$4 billion by 1990.

The US, which would be the prime contributor to such a fund, had already indicated that they found such a proposal unrealistic. Their opposition was based not so much on the size of the fund, but the way in which contributions would be assessed. Automatic contributions, they claimed, were incompatible with congressional feelings about the aid budget. If anything, extra funds should be strictly on a voluntary basis.

On Wednesday, the Group of 77 asked the developed countries directly whether they were prepared to support an overall increase in funds for science and technology for development. A meeting of ministers from the member countries of the European Economic Community took place on Thursday afternoon to consider the question — but failed to reach a consensus. At one end Britain's Minister for Overseas Development, Mr Neil Marten, argued that the recent cuts in Britain's public spending meant there was no money left for new aid purposes. At the other, the Danes, keen to build a progressive image among certain Third World constituencies, argued that a commitment to providing new funds could and should be made.

The EEC's failure to reach agreement had an unexpected effect. When it was announced, the Tunisian delegation, which is leading the negotiations for the Group of 77 and had been receiving some internal criticism for not being sufficiently tough, walked out of the committee room and, at a hastily arranged press conference, announced that they were leaving unless things moved a bit faster.

The next day, Friday, things did begin to move — but not entirely in the way that the Group of 77 would have wished. The EEC presented an answer to their question, saying merely that the member countries' budgets for technological development were growing in real terms, and that additional resources for science and technology "will be available subject to the priorities made in requests from the

developing countries". (The US had already said that they were prepared to agree to providing some extra funding).

Soon after the EEC reply, however, Sweden tabled a set of proposals suggesting among other things a very different mechanism than the one proposed by the Group of 77, in the hope that it might form the basis of a compromise proposal on which everyone at the conference might eventually agree.

Central to the package of proposals, which had been prepared by the United Nations Development Programme (see page 710), was the idea that the developed countries should commit themselves to contributing immediately to a more modest fund and, together with the developing countries, work on ways of increasing both contributions and uses of the money at a later date.

This proposal, which many had been expecting, but not quite so soon, put the Group of 77 in a difficult position. On the one hand, they had not been able to convince the developed countries that their own financing demands were reasonable. On the other, it was difficult for them to show that the Sweden/UNDP proposal was unreasonable. Yet it was already known that the US had indicated they would probably support the latter: and this itself was sufficient for the G77 to be sceptical.

The Group's reaction was to play for time. The US and the EEC, which both said they found the Swedish proposals "interesting", suggested that more details should be heard from UNDP administrator Mr Bradford Morse, who was already waiting in the wings. The Tunisian delegation, still under pressure from several of the more militant G77 countries, asked for time to consider the Swedish proposals prior to hearing Mr Morse — who subsequently agreed to reappear on Monday morning (returning to Paris in the mean-time).

The mere fact that the proposal was on the table, however, and that everyone knew it was possibly the only basis for a compromise agreement, meant that the whole initiative had passed from the Group of 77 to the developed countries. Indeed it became clear that some of the more powerful of the Group of 77, who had so far played a relatively back-seat role in the negotiations, did so partly because they were more in favour of the UNDP plan, but were reluctant to be seen splitting G77 solidarity.

By Monday, the whole tone of the conference was different from the previous week. No-one was pretending that the G77's demand on the financial commitments was likely to be accepted, indeed with the first working group, concentrating on methods for building up indigenous infrastructures and the transfer

of technology from developed to developing countries, progressing even slower than expected, it was becoming increasingly unlikely that the conference would produce a plan of action at all.

At the same time the central issue for discussion shifted — as the US had planned from the outset that it should — from the financial to the institutional aspects. The G77 had proposed that a new intergovernmental committee be established directly under the United Nations. The US, in contrast, argued strongly that the central co-ordinating mechanism for science and technology issues within the UN system remain under the aegis of the Economic and Social Council (ECOSOC).

On Friday the EEC, partly at the suggestion of the British delegation, who had earlier heard some harsh things spoken about their uncompromising stand on the financial issue, put forward a proposal which it was hoped would meet the demands of both sides. This was that an intergovernmental committee, open to all states who wished to attend, should be established by the General Assembly. But that it should report to the Assembly through ECOSOC (which, however, would not be allowed to do any more than add comments to the report).

At the time of going to press, it was still unclear how this issue was likely to be resolved. The EEC proposal had received support from the G77. But on Monday, the US announced that they were still adamant the committee should be appointed by ECOSOC, leading the G77 to issue a statement claiming that the UNCSTD conference had suffered "a serious setback because of the position of one major country."

The Group of 77 added that they hoped the situation would change, but that "the position at present does not bode well". In reply, the US issued a statement regretting "that some countries appear to have misinterpreted or misunderstood some parts of the US position." How much of this was pure negotiating tactics, diverting attention from more significant issues — like the role and status to be given to the secretariat which will service this committee, and whether it should be based in New York, Geneva, Vienna, or some third world country — remains to be seen.

By Monday, the two sides in the negotiations had come sufficiently close to make most observers optimistic that there would be no sudden breakdown in negotiations, and that there was sufficient interest on both sides to engineer a "successful" outcome to the conference. In the process, however, it was clear that most of the more controversial of the developing countries' demands had been slowly whittled away during the negotiating process.

David Dickson

NGOs clash at side-show meeting

The main idea behind the non-governmental organisation's (NGO) forum, according to its organiser Dr Karim Ahmed, is to bring together the many different types of NGO with an interest in science and technology for development — especially those “bored” lobbying delegates at UNCSTD. What he regards as most important, however, is not how the forum is conducted but what happens after it. “If we do not have a plan to follow the conference, all our efforts will have been a complete failure”, he told the opening session last week.

Just what should be achieved and how, however, were points of fierce disagreement throughout the first week. Some NGOs, critical that the forum had no plans for making a direct impact on the main UNCSTD meeting, argued that it should prepare a statement for UNCSTD. But during discussions early on in the week, that idea was dismissed by suggestions from the organisers that the preparation of such a statement would be a waste of time and that the NGOs with consultative status can lobby UNCSTD anyway.

As a result of those discussions — together with criticisms that there was no room within the forum programme for discussion of the two background papers prepared in advance — it was agreed to devote brief daily meetings and one whole day session to trying to formulate a ‘plan of action’ for the NGOs after the conference. The plan of action should come up with something “at least in terms of mechanisms”, said Ahmed. One idea might be a clearing-house or secretariat for managing and channelling funds.

Agreeing on a firm plan, though, is bound to be difficult because of the diversity of the NGO group. There are basically three types of NGO, says Ahmed: those representing the professional interests of scientists, mainly from developed countries, those representing development groups often working at the village level in developing countries, and those which are politically active in issues such as environment and disarmament who want to impress on the Third World the dangers of “non-benign” technology. “What do these three groups have in common? They have rather diverse interests,” says Ahmed.

The group mainly representing the professional interests of first world scientists has already broken away from the rest. Before the start of the forum, the Group of 19, which was formed last year and issued a ‘declaration’ at the Singapore Symposium, announced the setting up of a

Organiser criticised but not bowed

Ziauddin Sardar profiles Dr Karim Ahmed, chairman of the forum for non-governmental organisations

Next to Joao Frank da Costa, the Secretary-General of UNCSTD, Dr Karim Ahmed, the chairman of the NGO forum is probably the most hard working man in Vienna this week. While da Costa has received a considerable amount of attention, little is known about Ahmed who, after all, has just as great a responsibility.

Ahmed, a Pakistani ex-patriate, who became a naturalised American citizen last year, describes himself now as a “typical middle class American”. He traces his background to Japan, where his father, originally from Bombay, was an import-export merchant. He married a local girl, and just before the Second World War broke out, took his wife and children to Shanghai where they spent four years. In 1945, the family moved to India, and after the partition to Karachi.

In Karachi, Ahmed's father was responsible for setting up Habib Textile Mills, one of the major textile mills of Pakistan. Ahmed confesses that “the mill does not belong to the family any more” But he is “well off”, and the standing of his family in Pakistan is on a par with Adamjee, Dawud and other textile industrialists. In religious affiliation he belongs to the Memons, a pious and strong group of Sunni Muslims.

Ahmed did not follow his father into



Ahmed: interested in mobilising people.

new body with the aim of implementing the principles emerging from UNCSTD. The new body is to be based on the International Council of Scientific Unions (ICSU) and will seek funds from UN agencies for specific projects. Its precise nature is to be decided on by a preparatory meeting sometime this week.

Many of the other types of NGO view the Group of 19 with suspicion, seeing it as an elitist organisation made up of scientists who ask how their knowledge can be applied to development rather than taking the developing country approach of asking what are the problems first.

The main discussion at the NGO forum,

the textile industry. “My interests were more academic”, he says. He graduated from the University of Karachi and then decided to pursue his higher studies in the United States. After obtaining a masters in chemistry from the University of Minnesota, he moved to Harvard where he was awarded a doctorate in biochemistry in 1969.

During his student years, Ahmed developed a keen interest in radical issues. He worked for years as a volunteer for black organisations, and during the mid-sixties with Ralph Nader's public interest research group.

“I have always been conscious of social aspects of science and technology”, he says. His first major research project was a lead-poison screening programme which assessed the potential impact of lead poisoning on the black community. Ahmed says proudly that “it led to many city ordinances”. He now spends most of his time investigating the environmental impact of technology as a research director of the Natural Resources Defense Council, a cooperative venture between lawyers and scientists.

Ahmed's involvement with the UN began only two years ago. “I have always been interested in mobilising people” he says. He saw the NGO forum as a challenge and “went out” for its chairmanship.

There has however been considerable criticism of Ahmed's handling of the forum. He accepts the criticism as “symbolic”, saying “there are some people who must criticise”, and considers that most of it is based on a misunderstanding of the role of the NGO forum. “We are not a counter-conference to UNCSTD”, he says. “Neither are we a political group. The purpose of the forum is to allow as many NGOs as possible to meet and discuss the important issues”. But he hopes that some of them will organise joint ventures and carry them out after the conference. □

during the first week however, did not centre on the politics of formulating concrete proposals for action. Attempts were made to discuss some of the more pragmatic problems of development such as the role of energy and environment policy and the arms race in development. But even this part of the proceedings has gone badly. Many delegates have criticised the organisation of the conference which does not allow useful debate. If the NGOs are to bring pressure on governments then they should be using the forum to thrash out specific proposals and identify areas for special attention, they argue.

Judy Redfearn

India demands "sharp break" from current practices

In a strong statement that reflected many of the demands being made by other member countries of the Group of 77, the head of the Indian delegation to UNCSTD strongly criticised the way in which current forms of assistance from the developed world tends to increase the dependence of the developing world and to accentuate disparities in scientific and technological capabilities.

"A sharp break from this pattern so as to lead to a better sharing of the capabilities and to a lessening of global income disparities is called for" Dr D I Lakdawala, deputy chairman of India's Planning Commission, told the conference.

The experiences of two International Development Decades had led to a realisation among developing countries that the present structure of science and technology could not fulfil the types of promises that had been made earlier, such as providing growth with social justice and fulfilling minimum basic needs, he said.

Pointing out that a science and technology plan had been a central feature of Indian economic planning in recent years, Dr Lakdawala said that the Indian experience stressed, in particular, the need for a proper delivery and extension system to ensure that the fruits of science and technology became widely available.

"It also underlines the need for a proper infrastructure and appropriate economic policies and fiscal incentives to encourage application of scientific and technological discoveries on the farms and in the factories. And in this task developing countries can greatly benefit from mutual cooperation," he said.

In the past, a purely sectoral approach to issues such as technology transfer or access to information had often failed to develop endogenous capability in science and technology and to bring a proper share of benefits to developing countries.

It is essential that in future the selection of sectors and programmes and the methods of their execution must fully reflect the preferences of the developing countries concerned. The policy should be to build up full endogenous capabilities so that in the name of immediate benefit, dependence is not perpetuated and self-reliance is fostered."

It was along these lines that the countries of the Group of 77 had constructed the draft programme of action being debated by the conference, to which India had given full support. □

● India's efforts on primary health care — see page 716



Happy home-coming for cosmonauts Ryumin (left) and Lyakhov

Soviet spacemen return to Earth

Long-term orbital missions such as that completed recently by cosmonauts Vladimir Lyakhov and Valerii Ryumin, will be the mainstay of the Soviet manned space programme for the foreseeable future, according to a recent statement of Academician R.Z. Sagdeev, Director of the Institute of Space Studies. Priority will be given, he said, to establishing how long a cosmonaut can remain in orbit without damage to his health — an important preliminary to the establishment of the permanently-manned orbital stations predicted by Tsionkovskii.

Not surprisingly, the programme of the mission has placed considerable emphasis on physical and psychological health. Prophylactic measures ranged from the use of the Chibis pressure-suit to simulate the cardiovascular tone of normal gravity to the visit of pop-singer Valentina Ignateva to the mission control centre to sing to the crew. No medical surprises were expected from the mission; on a previous mission there was considerable excitement when the 120-day deadline was passed, since this is the lifespan of red corpuscles. According to one of the medical team, Dr Anatolii Egorov, this time it was simply a matter of collecting statistical data of the body's reaction to gradually extended periods of weightlessness. Owing to the crew's "exceptionally conscientious" attitude to exercise, their health remained excellent.

The experimental programmes of the mission took the now standard forms —

aserophysical and geophysical observations, "Zero-gravity technology" including a new aluminium-spraying experiment, and biological and genetic experiments. Towards the end of the mission, they assembled an orbiting radio-telescope, inaugurating, in *Pravda* words "a new stage in space research, that of extra-atmospheric astronomy".

Ironically, however, while the space-biologists seek ways to extend the length of manned missions, there are hints that certain planners would like to reduce the human involvement. In a radio interview, one of the designers of the radio-telescope, identified only by name and patronymic as Valentin Petrovich, said that in the future all such equipment would ultimately be automated, although "in the first stages of development . . . the role of man is very great". Likewise, Dr Konstantin Feoktistov, a frequent spokesman for the programme stressed that the majority of the crew's time is taken up with communications, routine housekeeping and maintenance, medical checks and the need to keep fit leaving little time to carry out the experiments which are, presumably, the main object of having a space station at all. "It follows that more use should be made of automation", he said, as far as the experiments are concerned, with the crew simply "correcting the programmes, repairing and servicing equipment, and so on".

Vera Rich

UK tram lovers book Shuttle space

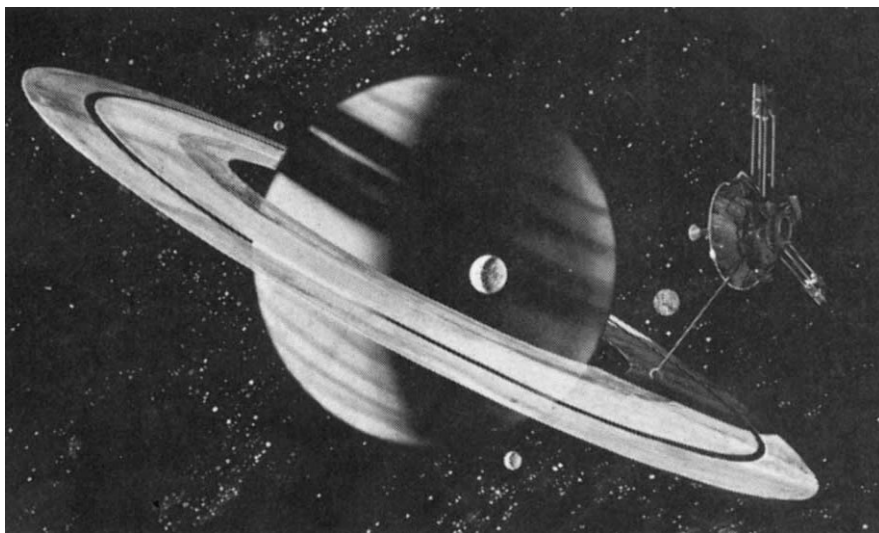
A call from the Tramway Museum in Crich, Derbyshire, has proved unusually fruitful for Professor Eric Laithwaite of Imperial College, London. Enthusiasts there have booked room on board a 1982 Space Shuttle flight and given their 1.5 cubic feet of space to Professor Laithwaite. "It's like a big birthday present", he said.

He will be designing an experiment to observe the behaviour of a precessing

gyroscope in zero gravity. This, he believes, may show that inertial mass and gravitational mass were not the same and would demonstrate an effect that could have enormous benefits in developing new propulsion systems.

The museum has booked the space on the Shuttle as a way of publicising itself and of celebrating the centenary of electric traction. □

Pioneer may collide with Saturn's 'E' ring



At 14.35 GMT tomorrow (1 September), the US Pioneer 11 spacecraft will cross the plane of Saturn's rings at an extremely shallow angle, and perhaps find out the hard way whether material in the planet's famous rings is spread out farther than appears on photographs taken from Earth. Because of the time taken for its radio signals to reach Earth, mission scientists waiting anxiously at Ames Research Center in California will not know for more than an hour whether the spacecraft has survived its 0.8 second passage through the ring plane; but the outcome will instantly resolve a long-standing debate over mission strategy — and at least a few private wagers.

There are three clearly visible rings surrounding Saturn: from the outermost inwards they are designated A, B and C. In 1969, some four years before the launch of Pioneer 11, Pierre Guerin discovered a fourth ring (D) lying even closer to the planet's surface. Mission scientists had originally voted 12 to 1 to have the spacecraft pass through the 'Guerin Division' between the C and D rings.

The decision, however, caused a conflict with scientists at the Jet Propulsion Laboratory, whose more sophisticated Voyager spacecraft will be flying by Saturn in November 1980 and August 1981. As its name implies, the Pioneer mission is a relatively low-budget scouting expedition, and JPL scientists wanted to make sure that the subsequent Voyagers would be following a safe path through potentially hostile territory. (Such lapses into the terminology of the Old West are understandable among scientists at Ames, where the laboratory facilities have been built on the site of an old stagecoach stop.)

The dispute was finally settled by officials at NASA headquarters in Washington, who decided to forego the advantages of having Pioneer pass through the Guerin Division only 14,400 km above the planet's surface. They estimated there would be too much chance of encountering

material within the division. Instead, they targeted the spacecraft to cross the ring plane at the presumably safe distance of 112,000 km, well outside the A ring which extends to only 77,400 km. After passing through the ring plane for the first time, Pioneer will approach within 21,400 km of Saturn's southern hemisphere, then cross the ring plane again at the same height.

Although this plan is still in effect, new data have revealed that even this 'outer option' may not be as safe as originally anticipated. Radar measurements indicate

By John Douglas, California

that there may be substantial material extending as far as 343,000 km from Saturn's surface, in a band now being tentatively called the E ring. This material does not appear strongly on ordinary photographs, probably because the E ring is composed of widely scattered large chunks, which can be more easily detected by radar.

When asked whether he thought the Pioneer 11 spacecraft would survive both its passages through the ring plane, chief mission scientist John Wolfe estimated the chances of survival at about 50%. If the spacecraft should hit a sizeable chunk of debris as it passed the rings it could well be destroyed, he said, but more than three-quarters of the experiments intended for the mission would already have been completed. In particular, Ames scientists would have photographs of Saturn with a resolution a factor of 10 better than can be obtained with the best Earth-based telescopes.

If Pioneer 11 does not survive its crossing of the ring plane, the greatest loss to the mission will be the proposed encounter with Saturn's principle moon. Titan, the largest known planetary satellite in the solar system, has a thick atmosphere, and there has been much speculation over whether it may be able to support life. Pioneer 11 is not instrumented adequately

to fully answer this question, but it has been designed to make some relevant measurements.

In particular, Pioneer instruments should be able to characterise Titan's atmosphere, by determining its opacity, and the composition and layering of its aerosols. Previous Earth-based observations have revealed that Titan's upper atmosphere contains an orange-coloured layer of smog, which could be due to organic compounds. Pioneer's measurements of particle bombardment will indicate whether methane in Titan's atmosphere is being dissociated in a way likely to form organic polymers. The presence of ammonia will also be sought, since nitrogen is a constituent of the reddish organic polymers thought to exist in the smog layer.

Finally, the spacecraft will measure the heat balance of Titan. The surface of the satellite is thought to be too cold (perhaps -193°C) for life to survive, but it has been suggested that there may be a warm layer in the atmosphere, caused by the greenhouse effect or volcanic action, where the evolution could have occurred.

At the very least, the Pioneer flyby of Saturn should produce some spectacular photographs of considerable scientific interest. Because of the spacecraft's speed, the rings will photograph only as white blurs, although other instruments on board may determine the size of the chunks of rocks and ice which constitute the rings. If Pioneer survives the passage through the ring plane, there should also be a startling visual effect as the rings suddenly reverse in brightness. During its approach, Pioneer will view the rings against the Sun, and they should appear dark, with the gaps bright because of scattering by the small amount of material resident there. After the ring plane passage, the rings will suddenly revert to their normal appearance.

Whatever happens during the crucial encounter with the ring plane, Pioneer will have lived up to its name, for JPL scientists will almost certainly use the findings of this preliminary mission to target the Voyager spacecraft. Meanwhile, Ames scientists have begun to argue that such pathfinder missions will continue to be valuable. Only four months ago, NASA headquarters announced that future planetary missions would be handled entirely by JPL; but since then the top administration of the agency has shifted and the new administration has stated that no laboratory will be automatically excluded from any future missions. It is now reconsidering Ames approach of low-budget exploratory missions as a viable option for future interplanetary exploration, particularly in view of the team's record in successfully controlling costs. The Saturn encounter may well provide the needed impetus for continuing this approach. □

news in brief

Watches warning: Stricter control of some digital displays for liquid crystal watches has been urged by the UK's National Radiological Protection Board following publication of its first consumer products report. The board began testing devices which could expose the public to radiation in 1976 and its report concentrates on two main devices — ionisation chamber smoke detectors and liquid crystal digital watches that contain gaseous tritium light sources. Most of the smoke detectors, were found to be satisfactory, and where problems were detected, the board reports that these have now been identified and solved. However, in some watches, the tritium light displays — which eliminate the need for batteries — were found to leak "significant" amounts of tritium when subjected to fairly violent tests. In some cases, this reached as much as 10,000 times the recommended limit for radiation emission. The board described this as "clearly unacceptable", although in reality it is unlikely that watches would be subject to the tests' stringent criteria. Nevertheless the board urged that greater care be taken in manufacture and added that in cooperation with the Federal Office of Public Health in Switzerland — where many of the light displays are made — a new set of guidelines was now being drawn up.

Mystery leak: Reports of a leak of radiation apparently emanating from the Israeli Embassy caused the London Fire Brigade's Chemical Incident Squad to go into action last week. When they reached the scene, however, the radiation had stopped. The radiation burst was picked up by monitoring apparatus on a fire appliance at the Old Court Place fire station next to the embassy. After two hours, it stopped, but resumed again for a short time later.

An embassy spokesman said they were "totally mystified" by the incident. There was nothing in the embassy that could be a source of radiation. The only possible culprit, a small X-ray machine used for examining suspect parcels, had been checked, and was found to be in good order. Fire Service headquarters at Lambeth were unable to comment on the type of radiation discovered. A Home Office scientist, however, told *Nature* that the only radiation-sensing devices carried on fire appliances were gamma-ray dosimeters and survey meters.

Helium supplies in danger: The US Department of Energy spends \$400 million a year on research into helium related technologies, and in hearings before the Subcommittee on Energy and Power of the Committee on Interstate and Foreign Commerce, helium was touted as the key to a "supertechnology" which could help mitigate the energy shortage through its use — as a liquid — to cool metals to superconductive state.

However, the supply of helium is in danger. Scientists said some 13 billion cubic feet of helium are released into the atmosphere each year as a waste product of natural gas extraction. The National Research Council estimates that at this rate, US supplies of helium (approximately 718 billion cubic feet) will last only another 30 years. The subcommittee wants Congress to pass the Helium Energy Act to enforce the conservation of helium.

Carcinogen data criticised: In an editorial in the *Journal of the American Medical Association*, the editor, Dr William Barclay, criticises scientific reports of carcinogens and the tests on which they are based. He accuses the media of encouraging "the chicken little phenomenon", heightening public anxiety by a constant barrage of news stories citing substances in general use as carcinogens, based on evidence which "is often tenuous and unconfirmed". Dr Barclay also castigates the methods of testing these substances, on the grounds that dosages used often exceed any that man might be exposed to. He also claims that the results of tests are finally evaluated by "persons of questionable expertise in the field of tumour histopathology".

UK chemical companies get drug warning: Chemical companies throughout Britain were being warned by the Royal Institute of Chemistry last week about some of their products that have been revealed as basic ingredients of a powerful hallucinogenic drug, known as BZ, which is said to be 100 times more potent than LSD. The move follows a campaign by the National Society for Crime Reduction and Social Justice — who are sponsored by the Church of Scientology — who have called for Government restrictions on the sale of the chemical ingredients of BZ, or 3-quinuclidinyl benzilate as it is properly known, and its inclusion within the Misuse of Drugs Act. The society says that the synthesis procedure for BZ is at present freely available from the British Patents Office and can be manufactured for less than £100. The group also wants the Ministry of Defence to release details of its experiments using BZ on soldier volunteers to provide public information about the drugs' lasting effects. They claim that BZ was originally developed in the US and tested as an incapacitating agent by the military and the CIA.

Pugwash council urges SALT ratification: The council of the Pugwash conferences has urged the rapid ratification of the SALT II nuclear weapons treaty by the US and Russia. "This step is essential for the avoidance of nuclear war and for enhanced international security", urges a prepared statement released after the council's recent meeting in Mexico City. Although it admits the treaty does not halt some strategic nuclear weapons now being developed, the council stresses that SALT II opens the door to further lowering of the level of nuclear confrontation and substantial cuts in nuclear weaponry. "In the interests of world peace, we urge the earliest ratification of SALT II and the immediate beginning of negotiations on SALT III", it concludes.



Three tears for the crocodile: The Indian government has launched a massive programme to save the crocodile from extinction, following an investigation by Dr H.R. Bustard of the Food and Agriculture Organisation into the declining population of the three varieties of Indian crocodile. One effect of this decline is a sharp rise in the population of predatory varieties of fish like the catfish which has led to a decreased yield of edible fish.

The Indian Government has therefore undertaken, on a priority basis, a project for crocodile breeding at several centres, including Hyderabad and Madras in the south and Orissa and West Bengal in the east. About a dozen officers are now being trained in the technique of capturing reptiles at the Central Crocodile Breeding and Management Training Institute, Hyderabad. When the hatchlings have grown to a length of about three feet, they are released into a reservoir. The Orissa Government has inaugurated a project for rearing gharial (above), one of the crocodile varieties, at the village of Tikerpada, with two villagers as guards. The nearby Sarktoshia Gorge area, of about 950 sq. km, has been declared a crocodile sanctuary. Another sanctuary is being set up at the Similipal tiger reserve, where again it is hoped to breed crocodiles in captivity.

The need for sanctuaries underlines the critical position facing Indian crocodiles. Quite apart from its skin, used for making shoes and fancy bags, crocodile meat is a valuable commodity, for Indian tribesmen believe it to be a cure for asthma.

from Javid Hasan in Bangalore

Barefoot doctors: symptom not cure

Anil Agarwal continues his assessment of Janata's effect on science in India

POSSIBLY the greatest achievements of the Janata government lay in its efforts to redirect attention towards rural health care. Soon after the government took over office in April, 1977, Mr Raj Narain, the then Janata Health Minister, announced a new policy for health care. An official report from his Ministry pointed out that despite the fact that India had built up a network of 5,300 primary health centres (PHCs) and 37,000 sub-centres in the rural areas, "when judged by the three simple yardsticks of literacy, life expectancy and infant mortality, our country comes low down in the 'Physical Quality of Life Index'."

Presenting his new policy, Mr Narain explained that the Janata government would for the first time try to "put people's health in people's hands". "My own conclusion," he said, "is that one of the principal reasons behind this (ill-health in villages) is inadequacy of people's participation in the health care programme. They have not shown the necessary enthusiasm for it. They have been almost indifferent."

The new government proposed that each village with a population of 1,000 would select its own community health worker (CHW). This person would be trained for three months in elementary health care, including traditional medical systems, during which he would receive a stipend of Rs200 per month. After his training, the CHW would work in his village and receive from the government Rs50 per month as an honorarium (not a salary) for his services to the community. He would also receive from the government medicines worth Rs50 every month which he would distribute free.

The CHW was thus expected to become, as Indian officialdom put it, a "people's health guard", a representative of the community on health matters, responsible only to his community. Though he would provide first aid and dispense a few medicines — a responsibility important in establishing his credence in the community — he would essentially be an agent for health education, inculcating better nutritional practices and mobilising the community on such issues as environmental sanitation, immunisation and family planning.

As such the new scheme was not based on any new ideas. The previous government had at various times appointed various committees to report on rural health care; and the Srivastava Committee appointed in 1974 had recommended rural health care schemes based on community participation. But few of the past Health Ministers, even the most erudite amongst them, ever showed enough courage to brush aside the objections of the powerful

medical establishment. Mr Raj Narain (often described as a "buffoon" by Indian political commentators for his rustic behaviour) rushed ahead with his scheme nevertheless. A senior Health Ministry official told me: "All he knew was that something had to be done to change the existing situation and it had to be done fast. He knew nothing about how it ought to be done. But he was not prepared to listen to the vested interests in the medical establishment either."

In just six months of Mr Narain's tenure, his Ministry launched the world's biggest scheme to train 'barefoot doctors' outside China. In the first phase of the scheme, villages under 777 primary health centres were chosen. The PHC doctors were expected to train community health workers in batches of about twenty. This phase ended in September 1978, and in the next phase some 1056 more PHCs were chosen. By December 1978 some 80,000 CHWs had been trained. The final aim is to train over half a million CHWs by the end of the current five year plan (1978-83) to serve each of the 560,000 villages of India.

Objections from the medical profession

The CHW scheme has attracted considerable criticism: some critics are hostile to its very aims, while the others are sympathetic to its aims, but doubtful about its achievements. The first type of criticism has come mainly from the powerful lobby of doctors. Professional organisations like the Indian Medical Association claim that the scheme will only legalise quackery. They feel that CHWs will do great harm — possibly even increasing drug resistance amongst disease organisms — by handing out medicines in inadequate dosages. In any case, they argue, India has a vast number of unemployed doctors and the government should first plan schemes to employ these.

The pressure to create more government jobs for such doctors is obviously very high, but the professional organisations seldom seem to realise that doctors are not prepared to work in villages — even if they were paid a salary ten times that of the CHW's honorarium. They have never suggested a sensible scheme for rural health care that would be within the ambit of the country's meagre resources. Positions for doctors in many PHCs remain vacant despite the high rate of unemployment.

Those who are sympathetic to the scheme point out several flaws that can be partly attributed to the speed with which it was organised. The training of CHWs has been left to PHC doctors who have in the past seldom shown any interest in rural health problems. These doctors are also expected to teach the CHWs about

traditional medical systems, about which they know nothing. CHWs are taught everything from Western medicine to Ayurveda and homoeopathy. But without a proper integration of Western and traditional medicine, the course has become a confusing pot-pourri. Not only does the CHW's training need to be improved, but research must also be conducted to standardise his technology.

But the more important questions that such sympathetic critics have raised draw attention to the highly stratified socio-economic nature of the Indian villages in which the CHWs have to operate. The idea of CHWs in India has clearly been helped by the success of their counterparts in China. But China has a different, and relatively egalitarian, social structure. Will the Indian CHWs actually serve those who need them most?

Even before the scheme began, Dr Debabar Banerji, Professor of Community Health at the Jawaharlal Nehru University in New Delhi, pointed out that it would benefit mainly the rich farmers who dominate the village councils. Though in theory it was an excellent idea to let the community choose its own health worker, in practice rich farmers would ensure that their own kith and kin were selected. Subsequent studies have confirmed this prediction. The poorest of the poor in Indian villages are the scheduled castes (untouchables). A survey conducted by Dr Ashish Bose of the Institute of Economic Growth, New Delhi, shows that in several PHCs the proportion of CHWs coming from scheduled castes is much lower than the proportion of the scheduled castes in the population as a whole.

Health planners also failed to appreciate at the outset how great an impact the high rate of unemployment in Indian villages would have on the selection process. Though the CHW honorarium was kept deliberately small to attract only motivated people, even this sum of money has aroused the interest of many unemployed young educated people, particularly sons and nephews of the dominant farmers. An extensive survey of 76 PHCs organised by the National Institute of Health and Family Welfare (NIHFW), revealed that as many as 8.5% of CHWs were graduates or post-graduates. Interviews with CHWs revealed that many of them think that the health scheme is basically a job-creation scheme, and that with experience they would probably be employed by the PHCs directly. Such individuals are certain to be disappointed soon, and are likely to move on to more profitable occupations when they can find them. Since the CHW scheme could be endangered if unsuitable candidates are appointed in large numbers, the Ministry of Health has now changed its



Indian village: "Low down in the Physical Quality of Life Index"

guidelines for the selection of CHWs, to include only candidates who have an occupation, are preferably above the age of 30, and are not highly educated.

The profile of the CHWs actually varies from one area to another depending on its socio-economic characteristics. In extremely underdeveloped hill regions like Meghalaya, where communal tribal structures still exist, the CHWs have been chosen not by village councils but by meetings of the entire village, and here a much larger proportion are women: as women seldom leave their village, they are expected to make good CHWs. In West Bengal, a highly politicised state with a Marxist government, the selection of CHWs has involved political parties. Several CHWs here are reported nevertheless to be particularly conscientious.

Dr Bose's survey of Punjab and Haryana shows that the green revolution prosperity, which has otherwise increased the power of the big landowners, has actually had beneficial effects on the CHW scheme. The first batches of CHWs in this region were certainly dominated by the nominees of the rich farmers who were attracted by the Rs200 per month stipend during training — a large amount even for this prosperous area — but they soon realised that their regular remuneration after training would be only the meagre

sum of Rs50 per month. The second batches have, therefore, been marked by a lack of response from the richer families.

This survey has also shown that the proportion of beneficiaries who come from scheduled castes is greater than the proportion of the scheduled castes in the total population. This is because the richer families in these states can afford private physicians and are therefore not interested in the medicines of the ill-trained CHWs. But in the poorer, more stratified, heartland of central India, it is more likely that both the appointments for CHW positions and the benefits of free medicine will go to the more privileged sections of the village community. A Ministry of Health report shows that in the relatively backward state of Uttar Pradesh, village leaders have taken an active interest in the selection of CHWs, often with very acrimonious results which have led to litigation. The NIHFV survey further reveals that over 20% of CHWs selected in this state were extremely young, between the ages of 15 and 19.

Another important question raised by sympathetic critics concerns the CHWs' role: are they supposed to be pill-pushers — some kind of second-grade doctors — or mobilisers of the community, encouraging and organising it to act on its own health problems? The aim of the Indian government was to put people's health in

people's hands through this scheme. Most primary health care pioneers like Zafarullah Chowdhury in Bangladesh, David Werner in Mexico, and Carroll Behrhorst in Guatemala, have also come to realise that health is determined by a complex set of socio-economic factors. Pills can do little to alleviate ill-health conditions that essentially arise from malnutrition and poverty. They have, therefore, slowly relegated curative medicine to a low priority and gone on to organise co-operatives to create productive employment, to establish grain banks for food security, and to ensure enforcement of land distribution laws.

Take scabies, for instance, a skin disease which is widely prevalent in India. The main cause of scabies is lack of personal hygiene, which in turn can be ascribed to a scarcity of firewood for heating water. A good health worker will, therefore, not just distribute an emulsion to treat the infection, but also educate people about the importance of cleanliness and mobilise them to organise fuelwood plantations.

Second-grade village technocrats?

But the Indian CHW is mainly a pill-pusher. The NIHFV survey shows that many members of the community, including the CHWs themselves see their main job as distributing pills, and most of the CHWs want to extend their role as minor village doctors by learning to give injections and use the stethoscope. Furthermore, the dominant social groups would certainly raise serious obstacles if any of them were ever to make any attempt to mobilise people. The Indian CHW is, therefore, unlikely to become a radical agent of change. He will, on the contrary, just become an extension of the medical bureaucracy, a kind of second-grade village technocrat, capable only of dumping bleaching powder in village wells to chlorinate and disinfect them instead of organising the community to observe better environmental sanitation. As the new scheme fails to address itself to any of the real and difficult issues that determine ill-health in Indian villages, it is the "same old medicine", according to Dr Banerji.

In fact, by organising such a programme of front line workers only for the villages, and not for the towns, the government is open to a charge of double standards, that it is setting different, and lower, standards of medical care for the rural poor — a charge, curiously enough, made most stridently by professional organisations, which are against the scheme completely.

But if it is accepted that Indian CHWs are going to be just pill-pushers, how good will they be at that job? One indication comes from the National Malaria Eradication Programme (NMEP), which has been using the CHWs extensively to distribute chloroquine tablets and collect blood slides. The NMEP's own malaria workers cover 10-15 villages with a total of

10,000 population. They are rarely able to visit a family more than once a fortnight, and considerable transmission of the disease can take place before they locate a case.

But the CHW is always present in the village. Dr S.M. Pattnayak, Director of NMEP, claims that in parts of Gujarat each CHW has reported 40 malaria cases a year through the blood slides he has collected, whereas the other health worker reported only 30. The CHWs also distribute chloroquine tablets as a presumptive treatment — anyone with fever first gets chloroquine. By launching this massive holding operation to control malaria, NMEP has been able to reduce the incidence from 6.4 million cases in 1976 to 4 million in 1978.

The Ministry of Health officials claim that CHWs have also been particularly useful in providing treatment for minor ailments like scabies, cough, fever and diarrhoea; in distributing eye ointments; in disinfecting drinking water wells by chlorination; and, in some areas, in motivating couples to use family planning methods. They turned out to be very useful in West Bengal and Bihar during the major floods of 1978, claim ministry officials.

Harmful side-effect of free medicine

Professor Banerji says that critics must appreciate even this limited usefulness of CHWs. According to him, the Indian CHW programme, despite its limitations, is "an even more pioneering effort than that of China. The Indian bureaucracy itself has provided an unprecedented opportunity to bypass the medical establishment, break its stranglehold, and reach the people directly. That became possible in China only in the post-revolutionary period. Here, this has happened in the pre-revolutionary period itself." Interestingly, Dr Banerji still remains an important critic of the scheme. His initial criticism had even prompted the Indian government to request the World Health Organisation to keep him off its expert committees.

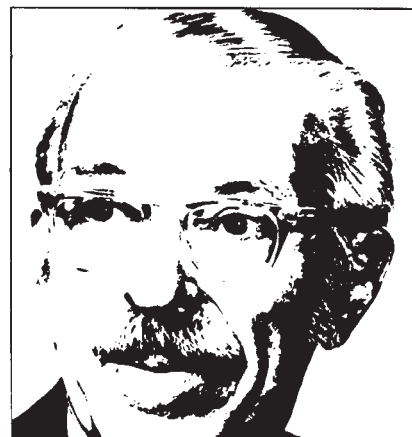
The future of the CHW scheme, nonetheless, remains open. Dr Bose fears that the CHWs' success as pill-pushers could prove to be their undoing. The more successful they are in distributing free medicines, the greater will be the demand for them. The government will then be expected to supply more free medicines, or else risk disappointment with the CHWs. But not even the richest government — least of all the poor Indian government — can afford to distribute free medicines to its entire population. CHWs will then have to get the community to contribute at least partially towards their own health care. But will they? The NIFHW survey shows that nearly two-thirds of the CHWs believe that the village councils will not be prepared to do so. The annual cost to the government of just maintaining the current meagre

Politics of popping pills

After each vitamin was discovered, the obvious task lay ahead of finding out how much of it was needed per person per day. These evaluations have continued through the years, and have included consideration of individual variation in needs. One would not expect the requirement to exceed greatly the amount of the vitamin obtainable from a diet containing a wide variety of typical foods. My early rule-of-thumb, in the 1940s, was that the daily requirement of a vitamin for a human being was roughly in the range of the amount needed per kilo of diet by young animals for maximum growth. For example, this amount is about 1.5 mg for thiamin and 2.5 mg for riboflavin. A percentage may be added to take care of a higher need by some people because of normal biological variability in populations.

Refined methods of measurement, including studies of blood levels, tissue saturation and excretion rates in human subjects are used by the Committee on Recommended Dietary Allowances of the Food and Nutrition Board, US National Research Council, in making recommendations. In its Report, issued every four years, the recommended allowances are defined as the levels of intake adequate to meet the known nutritional needs of practically all healthy persons. The allowances therefore exceed the needs of most people. The Report is useful and valuable as a condensed treatise on nutrition.

In recent years, self-dosage with vitamin pills has become a way of life for many American consumers. The practice is believed by some to produce 'super-health'. The head of the US National Cancer Institute was quoted in *Medical Tribune* recently as saying that he took 'a multiple vitamin preparation' to prevent cancer. No scientific justification was offered for this act of faith. Indeed, certain sceptics, including myself, think that the main effect of such dosage is to enrich urine with vitamins, and to line the pockets of vitamin sellers with dollars. The money involved has led to vitamin pills becoming a political issue. Legislation was specifically enacted by Congress in 1975 to prevent the FDA



THOMAS H. JUKES

from restraining the sales of over-the-counter high potency vitamin preparations.

High vitamin allowances were criticised recently in a book by Dr David Reuben, a physician whose biography lists his professional background exclusively as psychiatry. He described the National Academy of Sciences as 'just a nifty little private business, organized and owned by food manufacturers and vitamin sellers. It hires professors of nutrition to decide how much vitamins and minerals they think we should be gobbling and then pays them big salaries'. On 20 July, the Academy, unamused, filed a libel suit against Reuben and his publisher, Simon and Schuster. One wonders how Reuben came to be accepted by Simon and Schuster as an authority on the origin of recommended allowances for vitamins.

If indeed the Food and Nutrition Board had proposed overly *high* allowances, this would certainly have benefited vitamin manufacturers, who are currently enjoying the bonanza that has followed the high-vitamin-C boom. But the Board has been repeatedly accused by the 'health food' industry and other megavitamin enthusiasts of setting the allowances too *low* as a niggardly attempt to hold back the benefits that are alleged to result from high dosage with vitamins. How sad that the conscientious efforts of scientists should be denigrated by venal and incompetent critics! Gloomily, I await the fulfilment of the axiom that any publicity is better than no publicity, and the effect thereof on the sales of Reuben's output.

honoraria and supply of medicines will be substantial once the scheme is complete, amounting to over Rs650 million.

There will always be powerful lobbies to see even this expenditure reduced. In fact, if pressure on the central government to allow states greater autonomy in financial allocations succeeds, the CHW scheme could suddenly find itself in trouble, for many state governments remain lukewarm

towards it. A recent report in the *Times of India* quoted a senior Punjab government official as saying that his state would soon stop the CHW scheme. The statement was later denied, but this incident does show that financial decentralisation, though a good development in itself, could pose a threat. The end of the CHW scheme would indeed be unfortunate. As 'critic' Dr. Bose put it: "where is the alternative to it?" □

news and views

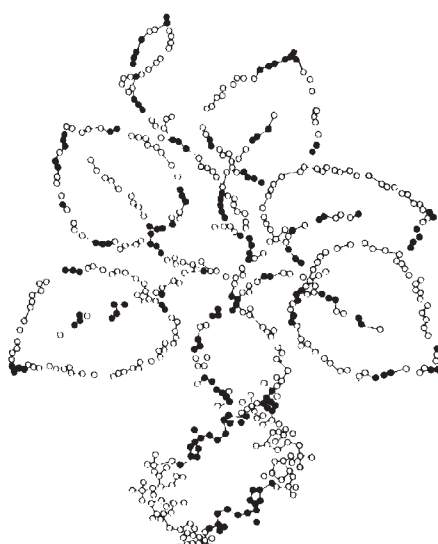
Plant molecular biology comes of age

from Wayne M. Becker

PLANT molecular biology is rapidly coming of age, or at least such was the consensus of 160 plant scientists from several dozen countries who attended a recent workshop on Genome Organisation and Expression in Plants*. The workshop underscored the number of highly qualified young investigators now active in this field, the technical elegance of much of the work, and the rapid progress being made in several areas of genomic analysis and expression in plants.

Probably the single most noticeable technical advance since the previous meeting of this group at Strasbourg in 1976 was the prominence of restriction analysis and cloning techniques. This was illustrated early in the conference by the elegant work of J. Bedbrook (Plant Breeding Institute, Cambridge) and colleagues on the cloning of DNAs characterised as highly repeated (heterochromatin from rye chromosome telomeres), repeated (rDNA from cereals) and unique (gene for the small subunit of ribulose-1,5-bisphosphate carboxylase). Additional applications of recombinant DNA technology were evident throughout the sessions, especially in the area of chloroplast DNA analysis.

The topic of nuclear genome organisation was introduced by W.F. Thompson (Stanford University) with a lucid discussion of sequence repetition and interspersions. To illustrate the spectrum of possible patterns, he considered in detail the genomes of two related species, mung bean (long period interspersions, with almost half of the single-copy DNA as sequences of 6,700 nucleotides or longer) and pea (short period interspersions, with repeated sequences scattered throughout the genome at intervals of about 200 nucleotides). In the same session, R. Flavell (Plant Breeding Institute, Cambridge) described detailed sequence analyses of several cereal genomes and suggested that much of the DNA of complex plant genomes originated by amplification and reamplification of compound units consisting of a unique or low-frequency



repeat sequence flanked by a common high-frequency repeat sequence.

Nuclear transcription was a topic ably addressed by J. Jendrisak and T.J. Guilfoyle (University of Minnesota). Jendrisak described the progressive appearance during germination of polymerase IIB at the expense of IIA, involving the cleavage of the 220,000 molecular weight subunit of IIA to the 180,000 subunit of IIB, and in wheat at least, the modification of a smaller subunit as well (from 25,000 to 27,000 molecular weight). Guilfoyle reported that the well-documented enhancement by auxin of nuclear transcription in soybean hypocotyls is due in part to a higher rate of synthesis of RNA polymerase I, but mainly to an increase in the propagation rate of this nucleolar enzyme. With respect to the actual expression of structural genes, R.B. Goldberg (University of California, Los Angeles) was able, by analysis of R_{0t} hybridisation kinetics, to distinguish between 'abundant' mRNA (70% of total, but only 500–1500 different kinds) and 'rare' mRNA (which contains the vast majority of different kinds of mRNAs, but in minute quantities). For the rare mRNAs, he estimated 60,000 species in the whole plant, and 25,000–30,000 in a given cell type, of which perhaps 6,000–10,000 are tissue-specific.

Hormonal regulation of gene expression figured in the report of L.S. Dure (University of Georgia), whose catalogue of mRNA subsets specific to various stages of cotton cotyledon embryogenesis and germination includes a novel set functional only during abscisic acid (ABA) arrest and therefore possibly ABA-induced. For gibberellic acid, T.H.D. Ho (University of Illinois) provided further support for a dual effect of the hormone on α -amylase induction in barley aleurone (transcriptional regulation during the first 12 hours, a possible translational effect thereafter). In the case of auxin, L.N. Vanderhoef (University of Illinois) insisted that the gene expression and wall loosening hypotheses are not incompatible, and tried again to arrange their early marriage. Taking advantage of the absence of fellow auxinologists who might have protested, he went on to declare the wall acidification model 'in trouble' and to propose that the primary mode of auxin at this level may be on Mg^{2+} uptake, with H^+ extrusion as a side effect.

As a prototype for the competent application of molecular techniques to a research programme already in progress, the work of K. Hahlbrock and colleagues (University of Freiburg) on the light-induced enzymes of flavonoid biosynthesis in cultured parsley cells was very encouraging. By purifying the enzymes, raising antibodies and assaying for mRNA activity both *in vitro* and *in vivo*, they could show rapid increases in mRNA activity due to light. Determination of enzyme half-lives allowed the light-induced enzyme appearance *in vivo* to be linked quantitatively to an enhanced mRNA availability.

The biosynthesis and genetic regulation of seed proteins was a topic that attracted special attention because of the implications for genetic engineering of crop plants, a possibility most convincingly advocated by T.C. Hall (University of Wisconsin), but tempered in the minds of many by an increasing awareness of the formidable genetic complexity of the

*A NATO/FEBS Workshop on Genome Organisation and Expression in Plants was held on 11–21 July in Edinburgh, and organised by Dr. C. J. Leaver, Department of Botany University of Edinburgh.

W.M. Becker is in the Department of Botany, University of Wisconsin - Madison.

storage proteins in most species. Good agreement on the general features of zein synthesis and mRNA properties emerged from the papers by B.A. Larkins (Purdue University), A. Viotti (Laboratorio Biosint. Veg., Milano), and B. Burr (Brookhaven National Laboratory), but controversy arose regarding the true level of heterogeneity of the maize storage proteins, the extent to which this represents expression of different genetic loci, and the number of copies of each zein gene per haploid genome. Probably the most impressive general progress on storage protein analysis in dicots was that reported by Hall on the G1 proteins of *Phaseolus vulgaris*. With the G1 mRNAs isolated and translated previously, they have now prepared and cloned cDNA for use as a probe in detecting genomic G1 DNA segments from a 'library' of cloned bean DNA. Hall also stressed the possible practical implications of differences in both time of onset and rate of protein accumulation and in the differences in total proportion of G1 protein between varieties (range: 18–70%).

Nitrogen fixation is another topic of great potential practical importance that is now increasingly amenable to molecular analysis. Working with *Klebsiella pneumoniae*, R.A. Dixon (University of Sussex) has been able to identify 15 *nif* genes, order them into 7 transcription units, and identify gene products for most of them. D.P.S. Verma (McGill University, Montreal) discussed proteins specific to nodulated plants, including leghaemoglobin (now established as a plant gene product, with about 40 gene copies per haploid genome) and a group of other nodule-specific proteins, including nodulin, a 35,000-dalton protein present in all nodules, but with no assigned functions as yet.

Ti plasmids

The Ti-plasmid of *Agrobacterium* emerged as a strong candidate for the 'most progress since Strasbourg' award. Attention now focuses on T-DNA, described by J. Schell (Max-Planck-Institut, Köln and Rijks-universiteit, Gent) as a continuous segment of the Ti-plasmid with a molecular weight of $10\text{--}15 \times 10^6$, since this is the DNA actually transferred to plant cells. T-DNA contains the genes for infection and for opine synthesis and consists of at least three functional units, only the central segment being conserved in all types of Ti-plasmids. Schell also discussed possible mechanisms for integration and replication of T-DNA and speculated on the Ti-plasmid as a vector for the introduction of selected genes into plant genomes. W.B. Gurley (University of Wisconsin) reported the presence of four major sites of T-DNA transcription in octopine-type tumour cells, one mapping in the region common to all Ti-plasmids. Whether T-DNA actually codes for structural gene products is a question rendered more amenable to

approach by the isolation and preparation of antibody against both octopine synthase and nopaline synthase, as described by J.D. Kemp (University of Wisconsin). Rounding out the Ti-plasmid story, M.D. Chilton (University of Washington, Seattle) proposed a model for genomic integration of T-DNA that depends upon the known homology between the fixed righthand boundary fragment of the T-DNA and the normal plant genome.

Chloroplasts and mitochondria

Organisation and expression of the chloroplast genome was a topic that attracted much attention and reflected in microcosm both the spectrum of techniques and approaches widely in evidence throughout the meeting and the progress during the 3 years since the Strasbourg conference. Using restriction mapping, hybridisation, and linked transcription-translation, L. Bogorad and colleagues (Harvard University) were able to map the following on the 85×10^6 circular DNA of the *Zea mays* chloroplast: two sets of ribosomal genes (as inverted repeats), one gene for the larger subunit of ribulose biphosphate carboxylase (colinear with the mRNA; no intervening sequences), and photogene 32, which codes for a 34,500 molecular weight polypeptide that is processed down to a 32,000 molecular weight thylakoid membrane protein. To this map, J. Weil (University Pasteur, Strasbourg) was able to add (in spinach) a minimum of 21 genes coding for tRNAs for 14 amino acids, T.A. Dyer (Plant Breeding Institute, Cambridge) ordered the genes for the RNAs of the large ribosomal subunit as 23S–4.5S–5S (5' to 3'), and reported sequencing of the region for 4.5S, 5S, and the 3' end of 23S rRNA. A scheme for rRNA synthesis in chloroplasts was proposed, involving the processing of a 2.7×10^6 rRNA precursor to 16S, 23S, and 4.5S rRNA, and the separate synthesis of 5S rRNA, possibly as a slightly larger precursor.

R.J. Ellis (University of Warwick) introduced the topic of transport and assembly of chloroplast proteins by tracing the fate of the carboxylase small subunit from its site of synthesis on cytoplasmic ribosomes through post-translational uptake and processing, possibly involving receptors in the chloroplast envelope that recognise some aspect of its tertiary structure. G.W. Schmidt (Rockefeller University) identified several cytoplasmically-synthesised chloroplast membrane proteins (CPII polypeptides 15 and 16) which are similarly synthesised as larger precursors, then transported into chloroplasts by a post-translational mechanism involving cleavage to their final length. The term 'transit peptide' was suggested to distinguish sequences involved in post-translational transport from the signal peptide responsible for cotranslation transport according to the now-classic Blobel model.

In a concluding session on the mitochondrial genome, C.J. Leaver (University of Edinburgh), presented results linking cytoplasmic male sterility in maize to the synthesis of a unique mitochondrial polypeptide with a molecular weight of 13,000 (or 17,500) for the T (or C) form of male sterility. In the case of T cytoplasm, nuclear restorer genes suppress synthesis of the variant polypeptide to an extent that correlates well with the sensitivity of the line to the toxin of *Helminthosporium maydis* race T, the fungus which attacks plants with T cytoplasm.

Clearly emphasised at the Workshop beyond the actual data and techniques was the need felt by plant molecular biologists for an international assembly of this sort at regular intervals. The triennial pattern set by Strasbourg (1976) and Edinburgh (1979) points toward another meeting in 1982, with some preference already expressed for the western shores of the Atlantic. □

Bone-marrow transplantation: man and mouse

by Miranda Robertson

APLASTIC anaemia, severe combined immunodeficiency and acute myelogenous or lymphoblastic leukaemia, particularly in adults, are all fatal diseases for which bone-marrow transplantation is the somewhat desperate remedy. Bone-marrow transplantation, in most of the centres where it is carried out, kills about half of those patients in whom the graft is successful. Death in these cases is caused by graft-versus-host disease — the immunological reaction of the grafted lymphocytes against the patient's tissues. In a smaller proportion of patients, the graft may simply fail. It was clear from clinical results reported at a recent meeting on bone-marrow transplantation* that while failures of engraftment are now fewer, the incidence of chronic graft-versus-host disease has increased. The meeting brought together clinicians, who work with random-bred men, and laboratory immunologists who work with inbred mice, in the hope that the clinicians might identify outstanding problems to which the research immunologists might supply the solutions.

In the case of aplastic anaemia and severe combined immunodeficiency (SCID), one fundamental problem is that clinicians do not fully understand the disease they are attempting to treat. This is reflected in some anomalies in the results of bone-marrow transplantation in such patients. For example, two of seven

*The International Titisee Conference on Critical Issues of Bone-marrow Transplantation was held in Titisee on May 10–12, 1979 and sponsored by Dr Karl Thomae GmbH.

surviving patients out of thirteen given transplants for aplastic anaemia at the Hammersmith Hospital eventually recovered full autologous bone-marrow function (E.C. Gordon-Smith, Hammersmith Hospital) which suggests that the function of at least some successful transplants in such patients may be to tide them over until they spontaneously recover, which some of them do on steroids alone. Exactly how high the proportion of spontaneous recoveries may be is not clear: of a series of 168 at the Fred Hutchinson Cancer Center in Seattle, only four spontaneously recovered. Whether in these cases the anaemia is caused by a toxic substance which clears only very slowly, or whether it is a temporary autoimmune problem is not known.

In SCID, on the other hand, it is quite common for successfully transplanted bone marrow to replace everything but B cells (R. O'Reilly, Sloan-Kettering, New York), which suggests a regulatory defect induced by the host. It is unclear at what stage bone-marrow differentiation fails in SCID (R. Parkman, Children's Hospital Medical Center, Boston) and this is hardly surprising since in spite of recent advances in mouse immunology, the ontogeny of the immune system is not understood.

A central problem is of course tissue matching, for which there is no entirely satisfactory test. A mismatch may lead either to the rejection of the graft by the host, where the host is immunocompetent and complete suppression is not achieved, or to graft-versus-host disease. B. Torok-Storb (Fred Hutchinson Cancer Center, Seattle) reported that the inhibition of donor cell proliferation by the patient's cells in allogeneic cultures could help to predict graft failure, while D. Schendel (Institute of Immunology, Munich) reported positive cytotoxicity tests between the cells of related donors who showed no serological evidence of mismatch.

An even more subtle source of graft rejection which is not detected by either of these tests is the ill-understood mismatch that will lead heterozygous F_1 mice to reject parental bone marrow though they will accept skin grafts. G. Cudkovic (State University of Buffalo School of Medicine) reported that cells from the F_1 mice which did not lyse parental-strain targets could nonetheless be shown to bind to them *in vitro*, and suggested that this might correlate with their cytotoxic potential. Whether such subtle effects play any significant part in marrow rejection in human patients is unclear.

Although on the whole the chances of success for bone-marrow transplants across MHC mismatches are low, there have been reports of some remarkable successes (O'Reilly & R. Storb, Fred Hutchinson Cancer Center, Seattle). This may be because some MHC antigens are more important than others: W. Elkins (University of Pennsylvania) finds that to minimise graft-versus-host disease in mice it is more important to match for I-region

antigens (the mouse equivalent of the human DR antigens) than for K or D antigens (the mouse equivalent of the human A and B antigens). The other possibility is that serological mismatches may not always be a reliable indicator of a strong T cell response in an individual case.

Given that perfect tissue matching will almost never be possible, and that mismatching of minor transplantation antigens will always cause graft-versus-host disease except in monozygotic twins, can anything be done to remove the cells that will react against the host before the bone marrow is transplanted? Much effort has gone into establishing the identity of the cells in question, and S. Theifelder (Institute of Haematology, GSF, Munich) and W. Müller-Ruchholz (Kiel University) have both contributed to their identification as T cells. J. Sprent (University of Pennsylvania) reported that anti-theta serum will prevent graft-versus-host disease across minor histocompatibility barriers in mice, which suggests that the T cell that mediates graft-versus-host disease must be present as a mature allo-reactive cell before exposure to the host antigens. The practical implication of this is that graft-versus-host disease could be largely prevented by the routine use of effective anti-theta antisera to remove such alloreactive cells before transplantation.

It has recently become fashionable to speculate that graft-versus-host disease may be the manifestation of a temporary failure of the donor lymphocytes to regulate one another's activities in a normal manner. There were several reports of abnormal immunoglobulin production during the early stages of graft take. Storb has seen peaks of nonspecific antibody in patients who do not respond normally to specific antigen; Parkman finds abnormally high levels of IgE; and L.J. Dooren (University of Leiden) has found that initially high levels of immunoglobulin subside when donor T cells become apparent. Parkman reported some preliminary evidence supporting the view that acute and chronic graft-versus-host disease are two different diseases, the acute form being a failure of suppression (Reinherz *et al.* *New Engl. J. Med.* **300**, 1061; 1979). Three patients with acute graft-versus-host disease had no measurable suppressor cells and the appearance of these cells coincided with the resolution of the disease. Patients with chronic graft-versus-host disease were more heterogeneous: two had no suppressor cells and high levels of polyclonal immunoglobulin; four others had two or three times the normal level of suppressor cells. It seems likely that neither diseases such as aplastic anaemia and SCID, nor the anomalous behaviour of lymphocytes derived from grafted bone marrow will be understood until more is known about the ontogeny and normal regulation of the immune system. Immune

Miranda Robertson is an Associate Editor of Nature.

ontogeny also bears crucially on the immune competence of patients with allogeneic bone-marrow transplants. The key to immune competence, at least for viral infections but probably for all T cell functions, is that T cells recognise antigen only in association with self MHC antigens. Experiments with thymus transplants in mice have shown that T cells learn this specificity to self-plus-antigen in the thymus; R.M. Zinkernagel (Scripps Clinic) has now extended these experiments however to show that while T cells are incapable of responding to antigens in association with MHC antigens they have not encountered during thymic maturation, there is an equally necessary post-thymic maturation step. Thus if a congenitally athymic (nude) mouse of haplotype k is supplied with a heterozygous $k \times d$ thymus, the host T lymphocytes are able to kill virus-infected cells of haplotype k but not haplotype d, even though both haplotypes are present in the thymus. However, if such a mouse is then thymectomised (to prevent further T cell maturation) and provided with spleen cells from a donor of haplotype d, its T cells become competent to kill infected cells of both k and d haplotypes. This shows that there is a second stage of T cell maturation which depends on the haplotype of the lymphohaemopoietic cells. For bone-marrow transplantation to restore full immunocompetence, therefore, donor and recipient must share at least one of either the A or the B HLA antigens, and one set of DR antigens. These conditions are nearly always fulfilled because of the need to avoid rejection or graft-versus-host disease, but Storb and O'Reilly both reported surprising exceptions. For example, O'Reilly claimed to have restored full immune function with fetal liver and thymus grafts in spite of the fact that the patient's T cells were histoincompatible with her macrophages and B cells. Theoretically it should not be possible for such histoincompatible cells to cooperate, and claims of transplant patients who flourish with a mixture of their own and their donors' lymphocytes are regarded by many with great scepticism. There are two possible explanations for such cases: either there are just enough matched cells in each subset to give immune competence, though not enough to be detected by the methods used; or there happens to be sufficient cross-reaction between the incompatible cells to enable them to cooperate. It is known that such cross-reactions exist.

Puzzles of this sort are inevitable since we do not yet understand lymphocyte interactions in inbred mice, let alone in men. But it is clear that, grim as the prospect is for anyone ill enough to undergo bone-marrow transplantation, it is not as grim as it used to be. Even so, the recent improvement in the prospects, in particular at some centres, owes more to patient and intelligent empiricism than to fundamental insights derived from basic research. □

Sperms in 'parthenogenetic' freshwater gastrotrichs

from Geoffrey Fryer

PARTHENOGENESIS has been adopted as a means of reproduction by several groups of freshwater animals, among which the Cladocera and Rotifera are the best known. Such reproduction has several advantages, one of them being that, in climates where active life is profitable for only part of a year, populations can build up very rapidly if every constituent individual is a potential mother. The rapid increase of some planktonic populations of Cladocera in spring in temperate latitudes is an example. Such parthenogenesis is generally facultative and is followed by a period of sexual reproduction which results in the formation of 'resting' eggs that often overwinter or tide the population over other adversities such as a period of desiccation. By such means the ecological advantages of parthenogenesis and the genetic advantages of sexual reproduction are effectively combined in what is often referred to as cyclical parthenogenesis. In some populations of, for example, the cladoceran *Daphnia*, parthenogenesis has come to dominate the reproductive process and sexual reproduction takes place only at long intervals. The logical outcome of such a trend is to become completely parthenogenetic and some populations of planktonic *Daphnia* have conceivably achieved this status. Among the Rotifera, a whole group, the Bdelloidea, is believed to contain only obligate parthenogens, males being unknown, and the same is true of several species of freshwater ostracods. Such reproduction by means of which only female offspring are produced by a process not involving fertilisation is known as thelytoky. Thelytokous animals present problems to the taxonomist, who finds it difficult or impossible to apply the species concept to them, while their existence is something of an embarrassment to the student of evolution.

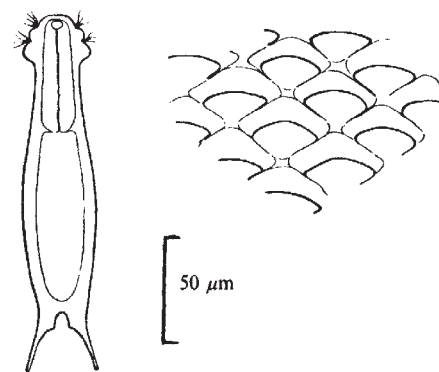
Another group of animals well represented in freshwater, and in that environment generally believed to include only obligate parthenogens, is the Gastrotricha. In the sea, however, some hermaphroditic forms are known. These microscopic animals — they seldom exceed 200 μm in length — sometimes regarded as comprising a distinct phylum and sometimes a class of the Aschelminthes, are common in a variety of habitats, usually associated with plants or mud surfaces over which they move by means of ventral cilia. Both quick-hatching (subitaneous) and resting eggs are produced but neither kind was believed to be fertilised, though in *Chaetonotus* two types of subitaneous egg

have been reported and the organism would not reproduce in culture unless two or more individuals were present, which hints at some sexual process. Now, following a tentative suggestion by Hummon that sperm heads may have been seen in *Lepidodermella squamata*, Weiss and Levy (*Science* 205, 302; 1979) have positively identified sperms in individuals of this species. These were grouped in packets, either singly or bilaterally, lateral to the intestine and seem not to be associated with any limiting membrane or ducts. Each packet contains only a few sperms — 16 is suggested as a possible maximum, a number that would correspond to the meiotic products of four spermatocytes. Individual sperms have also been found far removed from the packets but no signs of motility have been observed. Of the numerous individuals examined the proportion in which sperms were present never exceeded 8%. As ova are frequently present in sperm-carrying individuals these animals are not males but hermaphrodites.

In cultures, changes in the proportion of sperm-bearing individuals have been observed over short periods of time and the conclusion has been drawn that, as long assumed, parthenogenesis is indeed the usual means of reproduction but that the animals are really facultative hermaphrodites. It seems that hermaphrodites may appear only occasionally in nature, which may explain why they have been overlooked, and one now wonders whether, like the males of cladocerans and rotifers, they appear in response to certain conditions, and whether a kind of cyclical parthenogenesis occurs in some freshwater gastrotrichs.

Much remains to be learned about the sperms of *Lepidodermella*. It is not even known whether they actually fertilise the eggs, and although this has to be tentatively assumed it need not necessarily be the case. All kinds of peculiar cytological phenomena involving reproduction are displayed by small freshwater organisms — giant sperms in ostracods, resting eggs of *Daphnia* which, contrary to the general rule, can be produced without fertilisation, and so on — so caution is necessary. The small number of sperms produced and their lack of motility suggests internal fertilisation but it is not known whether self or cross-fertilisation occurs.

Provided cross-fertilisation is involved, the occurrence of a sexual phase, even if as a rare event, explains why one does not find a host of divergent 'strains' — a puzzle still presented by the apparently completely parthenogenetic bdelloid rotifers. It also makes it easier to understand speciation in freshwater gastrotrichs. There is now no



Lepidodermella squamata and details of its surface scales. (After Brunson).

need to invoke asexual speciation, at least in *Lepidodermella*, and this leaves the bdelloid rotifers — more than 200 species of them — as perhaps the only group in which this process seems likely to have occurred.

Thus the discovery of sperms in *Lepidodermella*, besides upsetting what is now seen to be a too easy generalisation, raises all sorts of interrelated problems of interest to the cytologist, ecologist, taxonomist and student of evolution, as well as to those who just like to understand how such minute creatures conduct their business. □

Squaring the circle

from Peter Lawrence

VERNON French's PhD thesis, an attempt to unravel the mysteries of insect leg regeneration, met with more success than he expected. His results revealed properties of insect legs that were similar to vertebrate limbs and in collaboration with Peter and Susan Bryant he launched the 'polar coordinate model' (*Science* 193, 969; 1976). In essence, this model attempts to explain formally why certain pieces of legs, insect imaginal disks and vertebrate limbs regenerate (remake the missing bits) and other pieces duplicate (make an extra set of the parts already present in the piece). For good measure the model also tackled a longstanding puzzle — the formation of supernumerary limbs after grafting. When a right leg is grafted onto a left stump two left appendages grow out from the graft/host junction, one on each side of the grafted right leg (Fig. 1).

The model draws liberally on the observations of Bohn who grafted pieces of cockroach legs together which had been cut at different levels in graft and host. Intercalary regeneration occurred at the host/graft junction and replaced the parts normally lying between. The model also

Geoffrey Fryer is Senior Principal Scientific Officer at the Windermere Laboratory of the Freshwater Biological Association.

Peter Lawrence is at the MRC Laboratory of Molecular Biology, Cambridge.

draws on gradient theory which has long been the preferred way of trying to understand how cells know their position. The essential point of gradient theory is that cells read the value of some continuous scalar variable, the cellular polarity being determined by the local maximum slope of that variable.

French, Bryant and Bryant proposed that this positional information is specified both circumferentially and proximodistally in the limb. When two pieces are placed together intercalary regeneration occurs to fill in the missing values, and this can occur both around the limb circumference and along its length. Experiments showed that the intercalary regeneration in the circumference always occurred by the shortest route — thus if $\frac{1}{4}$ th of the limb circumference is removed the limb regenerates; if $\frac{3}{4}$ ths is removed the remaining $\frac{1}{4}$ th duplicates.

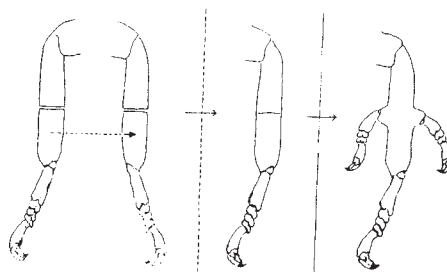


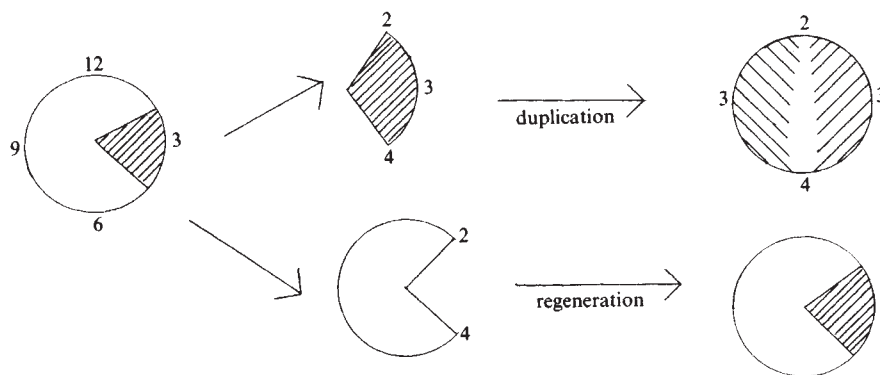
Fig.1 (From Bohn Wilhelm Roux Arch. 156, 449; 1965).

circle is clockwise or anticlockwise. French, Bryant and Bryant therefore proposed the 'complete circle rule' — that is, whenever a complete circle is formed proximodistal growth will occur to generate another appendage.

The flamboyant launching of this model (in both *Science*, and *Scientific American* 237, 66; 1977) has led to a flurry of experiment by others. It was soon found

briefly after wounding, but that soon after the cells became reallocated to different groups, each of which is then restricted to make structures in only one compartment. But when we specifically try to relate compartments — precisely defined parts of the pattern — to the polar coordinate model, we are somewhat lost.

Nevertheless a model that predicts much of the behaviour of pieces of disks and grafted limbs of both insects and vertebrates cannot be bad. We should now declare a moratorium on cutting and splicing experiments and stop tinkering with the model itself. I wish I knew what we should be doing, but one thing is clear; what started as a song and dance is in danger of becoming a song and dance routine. □



Reinhardt and coworkers (*Devl Biol* 60, 238; 1977) showed that the first step in this process was the healing together of the cut edges. This is then followed by the intercalation of missing values. It should be noted that after duplication the structure is mirror symmetrical and a further cut at right angles to the mirror plane would produce cut edges all at one level (3). After healing there should be no discrepancy and therefore no intercalation. Recently Littlefield and Bryant have tested this with the bilaterally symmetrical genital disk of *Drosophila* (*Devl Biol* 70, 127; 1979). They find, as expected, that cuts at right angles to the mirror plane produce a stable piece that heals but does not intercalate.

Another attraction of this model is that it formally explains the growth of supernumerary limbs after grafting, and whether they are left or right handed. This does require one extra rule. After grafting a distal piece of one limb onto a stump on the other side, circumferential intercalation is expected around most of the junction between host and graft — because of the varying disparity at the interface. However proximodistal intercalation is not expected (as the graft and host are cut at the same level). At two points on the circumference, intercalation will lead to the formation of a complete circle — a complete 'clockface' of numbers. It is at these points that the extra legs appear; whether they are left or right handed depending on whether the

that the complete circle rule does not always apply as incomplete limbs can and do regenerate (Slack & Savage *Nature* 271, 760; 1978). Theoreticians started to worry about how polar coordinates might be specified and whether two gradients at right angles might do just as well. In short, the model has been treated to a grilling in a field where models often get more attention than experimental facts. Of course, when formal models are found wanting a little tinkering can take account of the anomaly and the model can emerge unscathed with a shining new accessory.

As usual in developmental biology our understanding is fragmented and parochial, and our models reflect that. For example, the polar coordinate model is not concerned with the cell lineage of the regenerate or duplicate — whether the cells come from host or graft. Imaginal disks are known to be divided into two fundamental regions, 'compartments', each descending from a separate group of cells (Garcia-Bellido, Lawrence & Morata, *Sci. Am.* 241, 90; 1979). There are indications that limbs of cockroaches are also subdivided into two compartments. Unpublished experiments of Vernon French suggest that during intercalation the contribution of cells is limited to those structures in one compartment. Other experiments on imaginal disks (Szabad *et al. J. Embryol. exp. Morph.* 49, 229; 1979) suggest that compartment restrictions break down

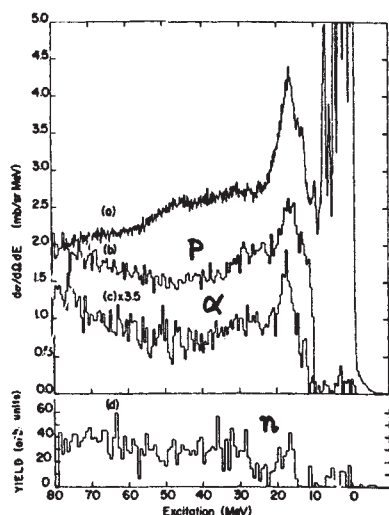
Alpha-particle decay of giant resonances in nickel reassessed

from P.E. Hodgson

THE recent work on the dominance of the alpha-particle decay mode in the decay of the giant quadrupole (E2) resonance in the nickel isotopes (see *News & Views* 280, 15; 1979) has stimulated several experimental and theoretical studies. Among these is a measurement of the probabilities of alpha-particle and proton decay by a more direct method, and this gives a different result that is more in line with experimental studies of other nuclei, and with theoretical work.

The new measurements were made by a group at the University of Maryland (Collins, Chang, Tabor, Wagner & Wu *Phys. Rev. Lett.* 42, 1440; 1979). They excited the giant quadrupole resonance by the inelastic scattering of alpha-particles, and determined the probabilities of emission of protons, alpha-particles and neutrons in coincidence with the inelastically scattered alpha-particles. Some of their results are shown in the figure. The top curve is the inelastic alpha-particle spectrum from the reaction $^{58}\text{Ni}(\alpha, \alpha')^{58}\text{Ni}$ at 140 MeV and at an angle of 16° to the incident beam. The giant quadrupole resonance is visible as a broad peak at an excitation energy of about 16 MeV in the target nucleus. The lower curves are the inelastic alpha-particle and a neutron, respectively. The accuracy was improved by integrating several coincidence spectra over angle, and then

P.E. Hodgson is a Lecturer in the Department of Nuclear Physics, University of Oxford.



The top curve is the total spectrum, and the three lower curves are of the alpha-particles in coincidence with protons, alpha-particles and neutrons respectively. The energy scale refers to the excitation energy in the ^{58}Ni nucleus.

subtracting the background continuum. The giant quadrupole resonance is visible in all the coincidence spectra, showing that the resonance decays by proton, alpha-particle and neutron emission.

These spectra were integrated to give the branching ratios for the decay of the giant quadrupole resonance, giving $59 \pm 12\%$ proton decay and $12 \pm 4\%$ alpha-particle decay. The difficulty of detecting the neutrons prevents the accurate determination of the neutron branching ratio. These results may be compared with the branching ratios from the statistical decay of a 2^+ state in ^{58}Ni at an excitation energy of 16 MeV. These were calculated using the Hauser-Feshbach theory, giving 51% proton decay and 6% alpha-particle decay. These are very similar to the observed values, suggesting that the giant quadrupole resonance decays statistically. To confirm this possibility it would be necessary to make similar measurements on a range of nuclei with different particle emission thresholds and penetrabilities, and to compare the results with statistical theory.

An investigation of the decay of the giant quadrupole resonance in ^{40}Ca by Youngblood *et al.* (*Phys. Rev. C* **15**, 246; 1977) gave $70 \pm 15\%$ proton decay and $21 \pm 10\%$ alpha-particle decay, quite similar to the values for ^{58}Ni . For this nucleus the statistical theory gives 74% and 21% respectively, in accord with the experimental results. The relative proportion of alpha decays increases towards the lighter nuclei, and the measurements of Knöpfle *et al.* (*Phys. Lett.* **74B**, 191; 1978) gives essentially 100% alpha decay for ^{16}O .

The new results for ^{58}Ni are thus in line with the trend expected from measurements of the decay of the giant quadrupole resonance in lighter nuclei. Detailed shell model calculations of the

In a recent comment (*News & Views* **280**, 15; 1979) P.E. Hodgson drew attention to an experiment at the National Bureau of Standards in Washington DC (Wolyneć, Dodge & Hayward *Phys. Rev. Lett.* **42**, 27; 1979) which appears to show that alpha particle emission is the dominant decay mode of the isoscalar giant quadrupole resonance in the nickel isotopes. This surprising result, derived from an analysis of electron and bremsstrahlung excitation function data is, in the light of recent work at three other laboratories, more controversial than implied by Wolyneć *et al.* Work in this laboratory (McGeorge *et al. Int. Conf. on Nuclear Physics with Electromagnetic Interactions*, Mainz, 1979) using a similar experimental procedure but a different approach to the analysis, and alpha scattering coincidence experiments at Maryland (Collins *et al. Bull. Am. Phys. Soc.* **23**, 506; 1979; *Phys. Rev. Lett.* **42**, 1440; 1979) and Heidelberg (Knöpfle *Int. Conf. on Nuclear Physics with Electromagnetic Interactions*, Mainz 1979) have not succeeded in reproducing the NBS result. We should like to point out that a number of unsubstantiated assumptions were made in the NBS analysis and that these taken with other considerations cast considerable doubt on the derived result.

Hodgson points out that the analysis depends on the accuracy of the E1 and E2 virtual photon spectra calculated from theory; there is no experimental test of the E2 spectra (Hayward *Int. Conf. on Nuclear Physics with Electromagnetic Interactions*, Mainz 1979) and tests of the E1 spectra (for example Martins, Wolyneć & Moscati *Phys. Rev. C* **16**, 613; 1977) do not seem to be better than 5%. The accuracy of the E1 spectrum is important because the E2 strength is essentially derived from the small difference between the calculated E1 contribution to the electro-alpha cross section and the experimental measurements.

The analysis also depends on knowledge of the monoenergetic photoalpha cross sections which are folded with the E1 virtual photon spectra to predict the E1 contribution to the electro-alpha yield. These cross sections have not been separately measured for nickel and the accuracy of the excitation function data

obtained so far does not allow determination of both the multipole contributions and the photo-alpha cross sections. It is therefore necessary to make some assumptions; in the NBS analysis an asymmetric Lorentz shape is assumed for the photo-alpha cross section. The resulting E1 contribution to the electro-alpha cross section increases less rapidly with energy than the experimental data and the difference could also be due to a high energy tail on the photo-alpha cross section.

We have performed a statistical model calculation of the photo-alpha cross sections starting with the photo-neutron cross sections measured by Berman (Lawrence Livermore Laboratory Report No. UCRL 78482, 1976) and find a larger high energy tail than the Lorentz shape of the NBS analysis. We then find that the electro-alpha data can be fitted adequately by including an E2 contribution which exhausts only 10% of the energy weighted sum rule. This E2 contribution is consistent with the alpha particle scattering coincidence experiments but much less than that found by Wolyneć *et al.*

Folding our calculated photo-alpha cross sections with the bremsstrahlung spectrum leads to 'radiator-in' yields 20-30% larger than those published by NBS but in agreement with recent measurements in our laboratory. In such experiments it is possible that enlargement of the beam spot size due to multiple scattering in the radiator can lead to counting losses unless a sufficiently large detector is used in the focal plane of the alpha particle spectrometer. In our measurements these losses are kept below 1%. Our electro-alpha results for which these losses are much smaller agree reasonably well with the NBS data.

The main point we wish to emphasise is that E2 strengths derived from electro-excitation function data depend sensitively on the virtual photon spectra and on the photo-alpha cross sections used in the analysis. It is therefore necessary to treat the results with considerable caution.

J.C. McGEORGE, A.G. FLOWERS, D. BRANFORD, C.H. ZIMMERMAN, *University of Edinburgh*.
R.O. OWENS, J.M. REID, *University of Glasgow*.

decay of the resonance in ^{16}O and ^{40}Ca reproduce this trend, thus confirming the experimental results.

It is thus no longer necessary to assume a particular alpha-particle structure to account for the decay of the giant

quadrupole resonance in the nickel isotopes. The earlier results that indicated a high probability of alpha-particle decay are thus probably attributable to unrecognised difficulties in the method used by the Maryland group. □



A hundred years ago

The British Association at Sheffield

The reports of the sectional proceedings in the local papers seem to us unusually meagre, though the space devoted to the Association both in provincial and in London papers becomes each year increasingly great. The *Times* alone this year has had several leaders on the Association generally, as well as on the principal addresses, and it is gratifying to

notice the decided improvement not only in the knowledge displayed in these articles, but also in the attitude of the leading paper towards science. In a somewhat hysterical article on Prof. Allman's address, the *Observer* of last Sunday seems to forget that science has all sorts of followers, and that the real workers rarely come before the public. Notwithstanding the apparently complete ignorance of the writer in the *Observer* as to what science really is, we cannot help agreeing with much that he says as to the present condition of the Association, and the urgent need of reform in its method of work, if it is not rapidly to degenerate into a "scientific garden party".

From *Nature* 20 28 August, 406; 18 79

Voyager 1 encounter with Jupiter and Io

The Voyager 1 encounter with the jovian system in March 1979 generated a wealth of new information about Jupiter's atmosphere, satellites, and magnetosphere. One of the most spectacular discoveries was the evidence of active volcanism on the satellite Io. Eight active volcanoes were discovered, indicating a level of activity higher than found on any other body in our Solar System.

This special issue of Nature presents a set of Voyager papers devoted to many of these discoveries. Principally, they cover the first detailed analyses of the jovian meteorology and the properties of Io. The Voyager papers are supplemented by Earth-based studies which relate directly to the spacecraft observations.

The results from the Voyager mission reported in this issue of Nature represent a significant increase in our knowledge of the jovian system. Frequently in the discussions, comparisons are made between features and processes in the jovian atmosphere, magnetosphere, and on Io with their terrestrial counterparts indicating the importance of comparative planetology. The success of the Voyager mission to Jupiter, indicated by these papers, will provide a new perspective for planetary studies.

Garry E. Hunt

Preliminary geological mapping of Io

Harold Masursky*, G. G. Schaber*, L. A. Soderblom* & R. G. Strom†

* United States Geological Survey, Flagstaff, Arizona 86001

† Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 86001

Terrain units and their global distribution on Io are summarised. A map of the complex region of Io is also presented.

VOYAGER 1 obtained moderate resolution (1–5 km per television line pair) images of about one-third of Io and images at lower resolution (5–20 km per television line pair) of an additional third of the surface¹. The low resolution images were obtained using several colour filters (0.35–0.6 μm) which discriminate a myriad of coloured surface units. The moderate resolution pictures are single-colour black and white and they show geomorphic features formed dominantly by volcanic processes². The absence of impact craters implies an extreme youthfulness for the surface that is consistent with the discovery of active volcanism on Io^{1,3–5}. This article presents a preliminary summary of terrain units and their global distribution on Io. The global map also provides a context for the accompanying articles that discuss topical aspects of Io's geology, including volcanic plumes⁴, the nature of volcanism implied from surface expression of landforms², curious layered terrains that are apparently eroded⁶, and a model for the volcanic eruptive processes⁷.

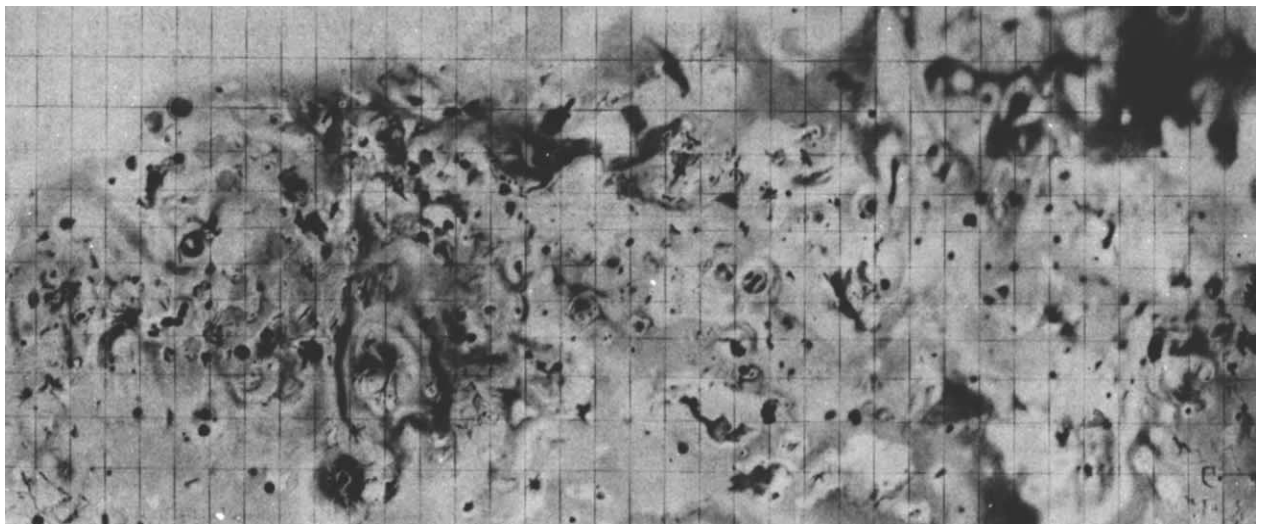
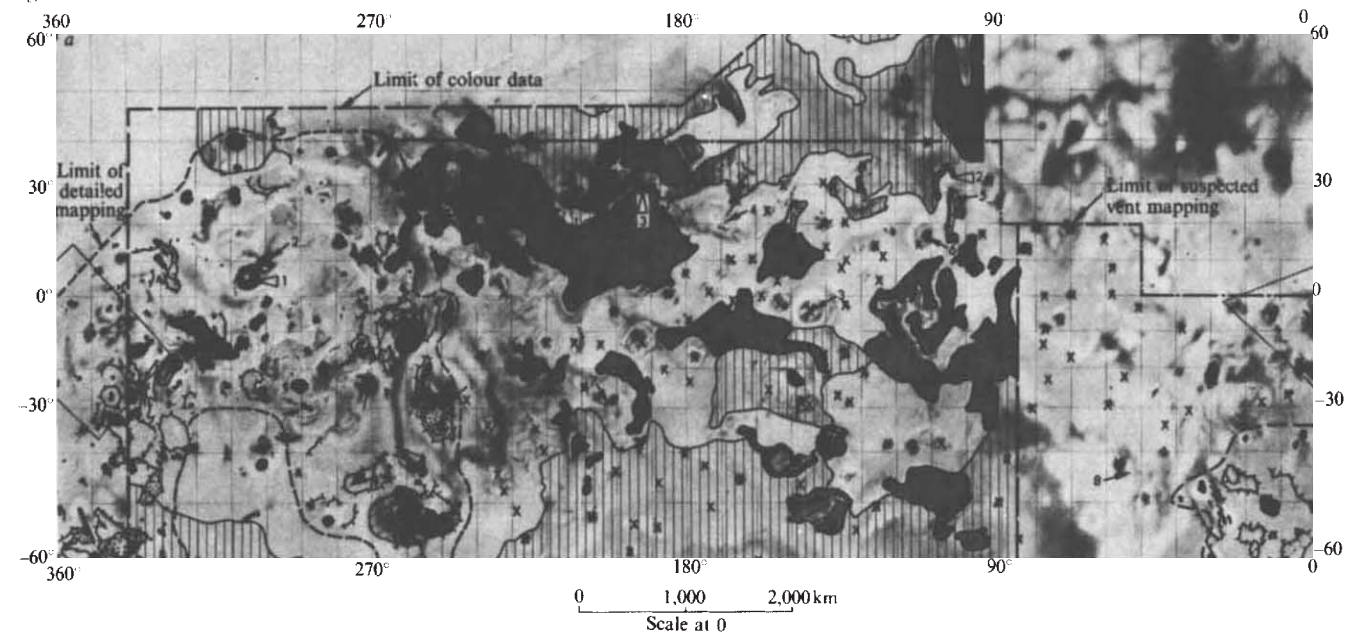
This paper also presents the first attempt to map units in a complex region of Io that shows interlayered flows, volcanic collapse pits, scarps, faults, and layered terrains. The maps were compiled using photogeological mapping techniques developed for mapping active terrestrial volcanic fields; the relative

chronology of surface units was determined from overlap and transection relations between surface units. A conventional planetary mapping technique that uses measurements of superimposed impact crater populations to establish relative chronologies for surface units could not be used because no impact craters have been identified on Io's very young surface. The detailed maps presented here (Plate 5a, p. 786) show that moderate resolutions obtained by Voyager 1 were sufficient to show such geological relations clearly.

Global map units

The most widespread terrain unit on the surface of Io is comprised of plains deposits that lie between the volcanic vents. In Fig. 1 these plains are divided into four general categories based on colour data where available (none between 0° and 80° longitude); the colour differences are also shown on the cover. The first plains unit consists of white to bluish-white regions of very bright material which may be surface deposits of SO₂ frost^{8–11} or the orthorhombic crystal form of sulphur¹². The second plains unit includes materials of intermediate albedo that show a variety of red, orange, and yellow hues; various forms of sulphur may be present on the surface in these areas as sublimates or sulphurous alteration^{1,12–14}. The third plains unit includes brownish hues (black and brown) that are limited

Fig. 1a



primarily to the polar areas. The fourth plains unit is very dark brown to black occurring as large diffuse deposits around a few vents.

These intervent plains generally appear smooth at moderate resolution but are in places characterised by extremely irregular margins with scarps of a few hundred metres or less (Fig. 1). These scarps, although globally distributed, show up well in the south polar region (Plate 3) and here show evidence of disrupted surfaces. Details of possible erosional processes are discussed briefly here and more fully by McCauley *et al.*⁶. Linear parallel sets of scarps that form graben, and linear features with no relief, have been mapped as normal faults. Protruding through the plains are isolated mountainous terrain (Plates 3 and 4, pp. 784, 785), most of which are estimated to be a few kilometres high based on limb profiles and shadow measurements near the terminator; the highest noted are <10 km.

The map of the planet (Fig. 1) shows the areal distribution of more than 300 suspected vents identified on moderate and low resolution photographs; the distribution of more than 150 vents identified on moderate resolution frames. Most vents mapped from the moderate resolution photographs resemble terrestrial calderas or pit craters. The floors (usually red and/or black) range from extremely shallow to ~1 km in depth. The 'calderas' in Fig. 1 have subcircular outlines, complex terraces, scalloped margins and radiating flows; 'collapse pits' have no discernible

ridges or surrounding deposits. Vents on the low resolution section of the map (×) are identified solely by their red or black circular forms or adjacent radiating patterns that closely resemble calderas and flows recognised in the moderate to high resolution photographs. The average distance between vents (>15 km diameter) in the moderate resolution coverage of Fig. 1 is ~200 km for both the equatorial and polar regions. The spacing between the inferred vents on Fig. 1 is ~300 km. The average vent size in the moderate resolution mapping areas is ~40 km; maximum vent sizes are ~250 km and ~100 km in the equatorial and polar regions respectively.

In the moderate resolution section of Fig. 1, volcanic flows with complex digitate patterns radiate outward from many vents; flows around other vents are broad and lobate to irregular in plan, with mottled, possibly coarse-textured surfaces². These flows are most common in the equatorial region; flows are rare in the south polar region (not shown on Fig. 1). Some calderas and pit craters are surrounded by bright and dark haloes (Plates 1, 2 and 4, pp. 782, 783, 785) (closely resembling dark halo craters on the Moon) and may be volcanoclastic deposits (volcanic ejecta comprised of silicate and/or sulphur compounds). Some elongate fissure vents were the sites of active plumes during the observation period⁴. High resolution images show complexly intersecting flows with low relief and volcanoclastic blankets in some areas. 43% in the moderate resolution areas from +30° lat

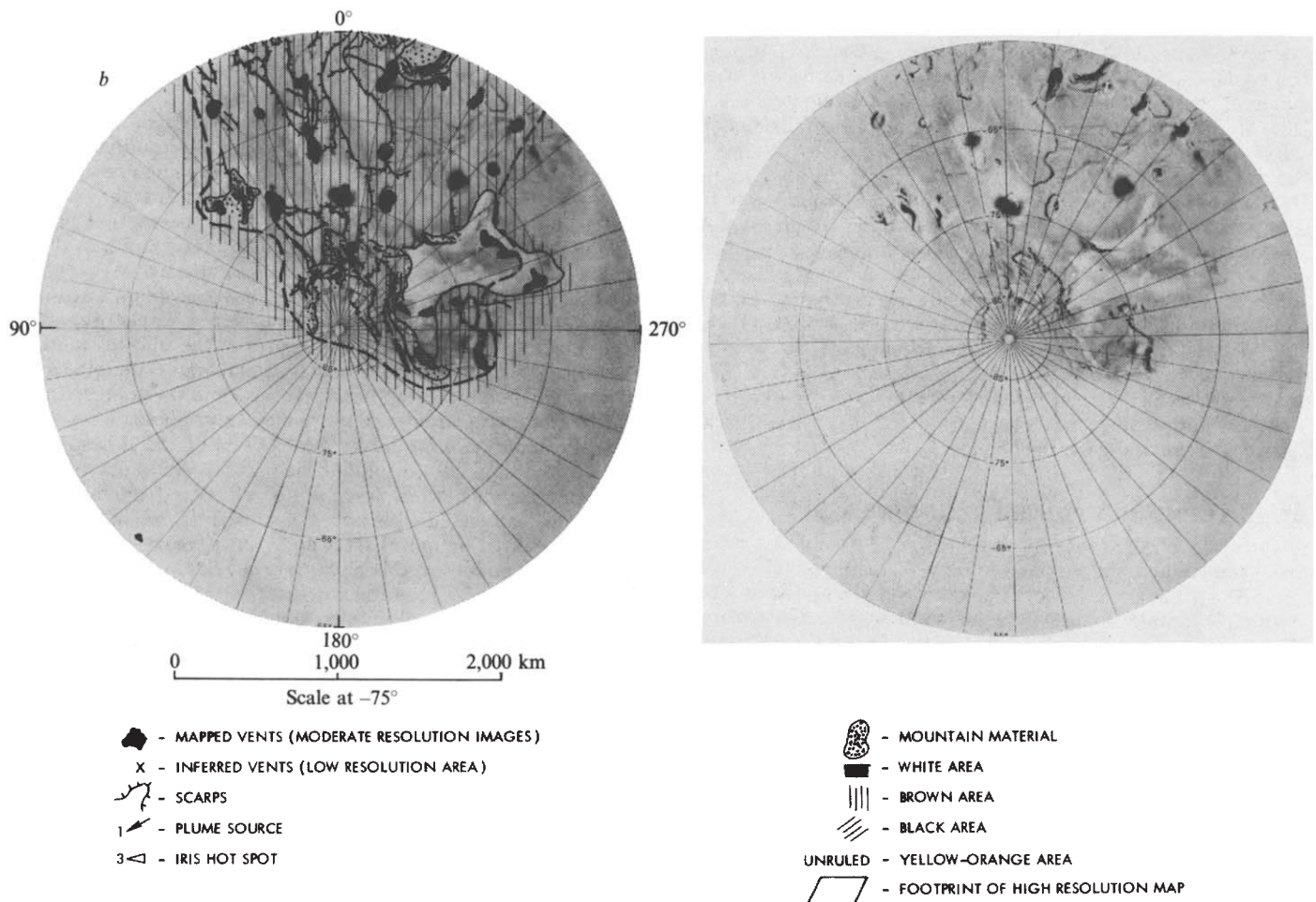


Fig. 1 Distribution of mapped and suspected volcanic vents in the *a*, equatorial; and *b*, south polar areas of Io; the base is an uncontrolled preliminary pictorial airbrush map of Io (mercator projection—equatorial; polar stereographic project—south pole by P. Bridges). Scale is at 0° latitude.

have mappable flow deposits while 23% of vents between -30° to -60° lat have such deposits. This percentage decreases rapidly towards the south pole to a maximum of 10%. 10–18% of the vents in the equatorial moderate resolution area have radiating deposits indicating possible raised topography; none of the vents mapped in the south polar region of Fig. 1 have such deposits. These percentages may be biased by the small sample area represented by the moderate resolution coverage.

Distribution and regional associations of global units

From Fig. 1 we can make observations regarding the distribution and interrelationships of colour units, terrain units, and surface landforms. All but one of the active volcanic vents, discussed in detail elsewhere⁴ are seen to lie within $\pm 30^\circ$ of the equator. Because half of the surface area of the body lies north and south of these parallels this seems to be a real and important observation. A statistical analysis was made to determine whether inactive centres are also distributed in any latitudinal fashion. The distribution of volcanic vents (calderas, collapse pits and radial flow centres) mapped in the moderate to high resolution area covered by Fig. 1 (60°S – 40°N ; 250° – 360°) was examined using 68 equal areas and modelling the expected number of areas containing 0, 1, 2, 3, 4 vents in the usual way. The χ^2 test indicated a probability $P=0.6$ that the computed² value would be exceeded by chance; the vents seem to be quite randomly distributed.

On the Mercator and stereographic maps (Fig. 1), regional fractures are notably lacking, only the local faults and scarps occur. We interpret the randomly distributed vents and lack of visible regional fractures to indicate that the vents have formed over widely distributed hot spots that are not controlled by strongly patterned convection cells that would lead to plates with detectable margins. This interpretation is consistent with the model of Peale and others that Io has a thin (<20 km), tidally heated crust¹⁵.

Isolated mountainous terrains in Fig. 1 are usually associated with calderas, collapse pits or active centres. Notable examples of mountains bordering calderas occur at $260^\circ\text{W } 10^\circ\text{S}$ (Plate 4a, p. 785), $255^\circ\text{W } 55^\circ\text{S}$, $350^\circ\text{W } 55^\circ\text{S}$ and $40^\circ\text{W } 77^\circ\text{S}$. The largest plume⁴ located at 250°W , 30°S is associated with two large tabular mountain units. The implications of these associations are discussed later.

Another important observation from Fig. 1 is the global distribution of colour units. As known from ground-based observations, the polar regions were slightly darker than the equatorial region^{16,17}. These units, classified as (dark) brown (to black plains) materials in Fig. 1 are found dominantly north and south of approximately 30 – 40° latitude. Where these materials bound the bright white areas and yellow-orange regions there is no discernible relief. The plains in the equatorial regions are either white or mixtures of reds and oranges. From the pattern of the distribution of the white areas shown in Fig. 1, we suggest that the entire region may have been a continuous blanket of white material and that the active centres, notably Plumes 3, 4,

5, 6 and 7 have the deposited materials which have buried or removed the whitish materials. Also, it is known from ground-based data¹⁸, that the hemisphere centred near 300° longitude is unusual compared with the rest of the planet. Figure 1 shows that there are few whitish areas in that region. Note also that the largest Plume (1) as well as a plume known to display an enormous UV enhancement (Plume 2), are located in this region⁴. We suggest, as in the case of the smaller plumes to the east, that these plumes have removed or buried the whitish materials. Another observation is that the complex surface flows occur dominantly in the orange and yellow regions. The absence of these complex flow units in the whitish materials as well as in the south polar darker units suggests two possibilities: (1) that the surface flows are, like the plumes, dominantly active in the equatorial region, or (2) that the whitish and dark units represent surface deposits (either condensates or enormous uniform flows) which have buried the complex flows characteristics of the orange-yellow materials.

High resolution volcanic relationships

Figure 1 shows the footprint of a high resolution image which has been studied and mapped in detail. This image (Plate 5a, p. 786) is situated in the orange-yellow plains and shows the extremely complex nature of the surface flows found in this unit. This image and accompanying map provide dramatic additional evidence of Io's complex volcanic nature. Numerous pit craters, shield caldera vents with diverse associated flows, pyroclastic and fumerolic deposits, and small volcanic cones have been mapped in this area. Two overlapping circular shield features are characterised by their high albedo summit calderas and sharply truncated outer margins. The circular marginal scarps appear to have been eroded; concentric faults associated with the features' central calderas may have controlled scarp formation and erosion. Dark and bright markings adjacent to small vents may be analogous to tuffs, sublimates or fumerolic deposits; they resemble deposits associated with currently erupting centres elsewhere on Io. Irregular markings on the flow surfaces may reflect variations in texture, chemical composition of the flows, physical state (for example, allotropes of S) or mantling surficial deposits^{1,2,12-14}.

The direction in which material flowed (indicated by arrows on Plate 5a, p. 787) gives at least some qualitative indication of local slope; the calderas on the summits of low shield-like constructs are thus associated with radiating, narrow, finger-like flows. There is little evidence of local, topographic slope for the pit crater deposits; eruptive materials appear to flow away from the crater in only one or two main directions. Materials from many volcanic vents overlap; thus a relative stratigraphy and chronology of eruptive episodes can probably be established; at least locally.

Possible geological models for the crust of Io

Our understanding of the geological character of Io's crust and the processes that dominate depends on the bulk composition of the major plains units exposed on the surface. On one extreme these plains may be thick accumulations of volatiles, that is, SO₂ and S₂ as discussed by Smith *et al.*⁷. On the other extreme, the plains might be predominantly silicate with minor amounts of sulphur and sulphur compounds producing the coloration. In that case the volcanism might be similar in style and character to the range of volcanic processes known in the inner Solar System.

Sulphur-plains model

In the first alternative, that is, thick accumulations of sulphur and sulphur compounds forming the plains, Smith *et al.*⁷ point out that assuming a thermal gradient consistent with the tidal heating model¹⁵ at a depth of a few hundred metres sulphur dioxide becomes liquid, and at a depth of about a kilometre sulphur itself becomes liquid. In this case topographic forms with relief or more than about a kilometre cannot be composed

predominantly of such volatiles as they would rapidly flow and sink back to a lower topographic relief. Thus, the mountains found scattered throughout the equatorial region and south polar region cannot belong to this volatile shell, but must be protuberances of silicates, presumably part of a silicate lithosphere. As Io's density (3.5 g cm⁻³) is very close to that of the Moon, it is likely that the interior of Io is dominantly silicate. The tidal heating model of Peale *et al.*¹⁵ suggests that the silicate crust could be of the order of 20 km thick, below that point liquid silicate would exist. Rough calculations by Smith *et al.* in starting with chondritic abundance suggest that a S₂/SO₂ outer shell is likely to be ≤ 10 km thick⁷. They also point out that below 4 or 5 km sulphur begins to vaporise and would probably be forced into solution or into the pores of the silicate mantle. Assuming that the dominant material making up the plains is sulphur and sulphur compounds, the depth to the solid silicate surface cannot apparently be more than several kilometres and then the mountains would be protrusions of the lithosphere through the liquid/solid sulphur-rich layers. Additionally, the numerous gas vents which form whitish effusions from the crust are usually associated with scarps of substantial relief⁶ (several hundred metres or more). McCauley *et al.*⁶ suggest that the liquid SO₂ zone is exposed on scarps of several hundred metres. Thus the liquid SO₂ aquifer has access to the surface and spring sapping, geysering, explosive erosion and simple gas effusion can explain the erosion along the scarps as well as the numerous occurrences of the whitish gas plumes.

Mountainous protuberances often ring volcanic centres, and this strongly suggests in the case of the two-layer model (that is, a silicate crust with oceans of volatiles with a solid crust of volatiles) that the apparent volcanic activity in the sulphur-rich layer may be excited by and localised by volcanic processes in the subjacent silicate crust. We suggest that the rings of mountains may in fact represent volcanic centres, perhaps calderas themselves in the silicate crust over magma chambers and that the collapse depressions are simply localised by 'sub-oceanic' volcanism ('oceanic' applies to the oceans of sulphur). Because of the density contrast between sulphur and silicates (nearly a factor of 2), similar to the density contrast between silicates and water by an order of magnitude, silicates that erupted from the protuberances would simply run down into the floors of the sulphur oceans. Volcanic constructs on the floors of the sulphur oceans could excite sulphurous activity above them and perhaps build up to heights of several kilometres required to protrude above the sulphur layer. The fact that most of the silicate protuberances are only 2–4 km above the plains places constraints on both the density contrast as well as overall depths of the silicate magma chambers. If the silicate crust were many kilometres deep, then the construction might be expected to be substantially higher. The implied confinement of the silicate magma chambers to a fairly narrow depth region is consistent with the Peale *et al.*¹⁵ heating model that the outer shell of only 10 km of Io's silicate mantle may be solid.

Silicate volcanism model

The appearance of a wide variety of volcanic forms that are usually associated with chemically differentiated volcanic rocks implies that the eruptive rocks may be normal silicate magmas enriched in sulphur. The forms vary from those associated with mafic rocks, such as basalts (Plate 2a, p. 783), to intermediate rocks that form lobate flows (Plate 2b, p. 783) or low domes (Plate 5a, p. 786), to silicic rocks and ash flow tuffs (Plate 4a, p. 785) of latite or rhyolitic composition. Also the mountainous terrain (Plate 4a, p. 785) that is 4–8 km high must be of silicate composition as sulphur would flow like a glacier under these stresses. The thin silicic crust would still hold, but sulphur segregation would have taken place only to the extent that the eruptive deposits are sulphur rich, highly coloured, and local accumulations of sulphur deposits might have occurred. More intense study of the Voyager data may allow a choice to be made about the composition (all sulphur or sulphur enriched silicates).

Time scales and surface ages

The absence of impact craters prompted Johnson *et al.*⁵ to estimate that the minimum resurfacing rate on Io is $>10^{-2}$ cm yr⁻¹. Resurfacing, meaning removal of craters, can be caused by various surface processes including volcanic flows^{2,12}, erosion⁶ and deposition from eruptive plumes⁴. Johnson *et al.* estimate that deposition from active plumes is 10^{-4} cm yr⁻¹. If this were an average rate for the planet then the entire surface would be coated by ~ 1 mm in, on average, 1,000 yr. Thus the multi-coloured surface flows are likely to be much younger than this. The processes that are erasing craters would also erase small calderas, scarps, and graben. From the resurfacing rate required to destroy craters ($>10^{-2}$ cm yr⁻¹)⁴ we can estimate that small calderas and scarp features, having relief of a few hundred metres to a kilometre, are <1 Myr in age; how much less we cannot be certain. The only candidates for ancient landforms are the silicate mountains. Although there are no impact craters on these mountains it is possible that such craters may have been erased by continually renewed mantles of sulphur and sulphur compounds; these mountains could conceivably be billions of years old. The surface colour patterns and landforms in the plains must be less than 10^3 and 10^6 yr old respectively.

Received 16 July; accepted 3 August 1979.

1. Smith, B. A. *et al.* *Science* **204**, 951–972 (1979).
2. Carr, M. H., Masursky, H., Strom, R. G. & Terrile, R. *Nature* **280**, 729–733 (1979).
3. Morabito, L. *et al.* *Science*, **204**, 972–976 (1979).
4. Strom, R., Terrile, R., Masursky, H. & Hansen, C. *Nature* **280**, 733–736 (1979).
5. Johnson, T. *et al.* *Nature* **280**, 746–750 (1979).
6. McCauley, J. F. *et al.* *Nature* **280**, 736–738 (1979).
7. Smith, B. A. *et al.* *Nature* **280**, 738–743 (1979).
8. Fanale, F., Brown, R. H., Cruikshank, D. P. & Clark, R. N. *Nature* **280**, 761–763 (1979).

Conclusions

First, the distribution of mappable volcanic vents (all types) is statistically random. This hypothesis is consistent with a model assuming very thin crust, high degree of internal melting and a wide distribution of 'mantle hot spots' or convection cells; second a wide variety of volcanic vent types and a similar diversity of deposits exist which indicate a large variation in eruptive mechanisms, rates of effusion and possibly composition of silicate melts and/or volatile components such as sulphur; third the large expanses of layered deposit plains on Io seem to be eroded remnants of large scale volatile or volatile-rich ash flow deposits erupted from calderas, the relics of which may be the rugged, layered and elevated mountain materials.

The images of Io obtained by Voyager 1 are sufficient quality to provide important geological information on this unique body in our Solar System. Geological mapping on high resolution and controlled low resolution maps should provide detailed information on the mechanisms and chronology of eruptive events.

This work was supported by the NASA Voyager Project. We thank Art Dial, Jim Prosser, Mimi Strobell and Rick Terrile for assistance.

9. Smythe, W. *et al.* *Nature* **280**, 766 (1979).
10. Kumar, S. *Nature* **280**, 758–760 (1979).
11. Pearl, J. *et al.* *Nature* **280**, 755–758 (1979).
12. Sagan, C. *Nature* **280**, 750–753 (1979).
13. Nash, D. & Fanale, F. *Icarus* **31**, 40 (1977).
14. Nelson, R. & Hapke, B. *Icarus* **36**, 304 (1978).
15. Peale, S. *et al.* *Science* **203**, 892 (1979).
16. Murray, J. *Icarus* **25**, 233 (1975).
17. Menton, R. B. *Commun. Lunar Plan. Lab.* **10**, 35 (1973).
18. Nash, D. B. *Icarus* (in the press).

Volcanic features of Io

M. H. Carr*, H. Masursky†, R. G. Strom‡ & R. J. Terrile§

* US Geobiological Survey, Menlo Park, California 90425

† US Geological Survey, Flagstaff, Arizona 86001

‡ University of Arizona, Tucson, Arizona 85721

§ Jet Propulsion Laboratory, Pasadena, California 91103

Volcanic activity is apparently higher on Io than on any other body in the Solar System. Its volcanic landforms can be compared with features on Earth to indicate the type of volcanism present on Io.

AN exciting event of the Voyager mission was the detection of a diffuse, umbrella-shaped plume extending 280 km above the bright limb of Io¹. This discovery was followed by detection of at least seven other plumes² and several areas of anomalously high surface temperatures³, indicating rates of volcanic activity very much higher than on Earth, or on any other body so far observed in the Solar System. The surface of Io shows complementary indications of high eruption rates. The apparent absence of impact craters larger than 1 km in diameter indicates that the satellite is rapidly resurfaced and few, if any, surfaces older than 10^6 – 10^7 yr survive^{2,4}. Numerous caldera-like depressions, commonly surrounded by flow fields, suggest that volcanism is the prime resurfacing process. So numerous are the calderas that nearly all are within 300 km of another. We describe here the various volcanic landforms, compare them with similar features on the Earth, and make inferences concerning the type of volcanism on Io.

An important aspect of ionian volcanism is the apparent part played by sulphur. Ionised sulphur has been detected in the vicinity of Io's orbit from telescopic observations⁵ and by the

UV spectrometer on board Voyager⁶. Sulphur is also a likely cause of the colour variations observed at the surface⁷. Smith *et al.*² suggested that sulphur or sulphur compounds (H₂S, SO₂) may drive the plumes. Later Johnson *et al.*⁴ and Smith *et al.*⁸ argued that the less volatile gases, SO₂ and/or S_x, are more likely candidates than H₂S. This conclusion is reinforced by the discovery of SO₂ ice on the surface^{9–12} and the detection of SO₂ gas near one of the active plumes by the Voyager IR interferometer¹³. Most of the sulphur in the plumes probably returns to the surface on near ballistic trajectories. A small fraction, however, must escape to form Io's torus either by some thermal mechanism, aided by the volcanism, or by magnetospheric sweeping⁴. Because of the apparent importance of sulphur in driving the volcanic plumes, we give special attention to the possible role of sulphur in creating the volcanic landforms.

Although Io was systematically photographed as the Voyager spacecraft approached Jupiter, only a fraction of the satellite's surface was covered at the highest resolution. The best coverage is between 250° W and 30° W south of approximately 10° N. Most of this area was photographed with resolutions down to 0.5 km per pixel. Outside this region, discrimination of surface detail is difficult and interpretation is largely based on comparison with the better photographed area. The general descriptions given here may therefore not apply to the entire satellite. In particular, some of the brighter regions of Io, around 50° W, were not well-photographed and may differ in physiography as well as in albedo.

Description

Calderas: According to Williams¹⁴, "calderas are large volcanic depressions more or less circular or cirque-like in form, the diameters of which are many times that of the included vent or vents" and most are produced by engulfment. Calderas occur in every region of Io so far photographed. They are generally recognisable by their strong resemblance to terrestrial and martian calderas. In many cases, however, no relief can be detected and a caldera is inferred simply from the presence of a dark circular feature. The calderas are remarkably numerous, with more than five larger than 20 km in diameter for every 10^6 km², or over 200 larger than 20 km on the entire satellite. By comparison, the Earth, with 3.5 times the land area, has approximately 15 calderas larger than 20 km in diameter¹⁵. Over 5% of the ionian surface seems to be part of a caldera, either dormant or active. Where relief is discernible, the calderas are recognised as rimless depressions with steep, inward-facing scarps and relatively flat floors. Rubbly rims that characterise impact craters are absent. Although most are nearly circular, they range widely in shape; some have scalloped walls suggesting collapse about different centres, others have rectilinear outlines, and others have elongate, slot-like shapes. While most seem to have little internal relief, this may be deceptive since the viewing conditions for most of the available pictures are not ideal for discerning topography. In the highest resolution pictures, low escarpments and irregular ridges can be seen on some of the floors. Some, such as the 25 km diameter caldera at 52° S, 33° W, have floors at different levels like many terrestrial and martian calderas. Differences in elevation of the floor may also be indicated by sharply defined dark areas which are probably caused by puddling of lava.

One of the most striking aspects of the Io calderas is the associated albedo patterns (Plate 3, p. 784). The floors of most are very dark and the low reflectivity of many is accentuated by bright haloes around the crater. The combination of dark floors and light haloes gives Io its characteristic spotted appearance in the far encounter pictures. Patterns are discernible on many of the dark floors. For example, a faint concentric banding can be seen in the 35-km diameter caldera at 23° N, 227° W and the 200-km diameter, irregular dark feature at 5° N, 305° W (Plate 1, p. 782) contains a 60-km diameter bright 'island' and numerous bright spots within the dark area. Although most calderas are darker than their surroundings, some are brighter as at 28° S, 338° W and at 62° S, 20° W, and some have an albedo almost identical to their surroundings as at 74° S, 340° W and 50° S, 12° W. Many calderas have complex albedo patterns in which sharply demarcated, irregular dark areas contrast strongly with the rest of the floor. The dark areas commonly terminate abruptly against the caldera walls. Diffuse dark and light markings are also common.

Very few of the Io calderas seem to be associated with an edifice of significant vertical relief. The low relief seems to be real and not simply an artefact of illumination and viewing conditions since even calderas seen near the terminator show little evidence of an edifice. Some vertical relief must exist, however, as flows seem to radiate away from most of the calderas. The ionian calderas are thus similar to such large terrestrial calderas as Valles, New Mexico, and the Aso and Aira calderas of Japan which have little vertical relief and are associated with vents that produced large volumes of ash. They are distinctly different from the Hawaiian calderas and those of the Tharsis Montes on Mars, which are located at the summit of large edifices built mainly of lava.

Flow patterns: The pattern of flows around the calderas is difficult to discern because the topography of the flows can rarely be distinguished and the albedo patterns are complex. Both dark and light flow-like markings seem to be superimposed on a surface which itself has distinct markings. Furthermore, the presumed flows in some cases seem to have altered the albedo of the former surface. The plumes cause further complication by masking the surface markings with diffuse dark and light overlays. However, certain patterns can still be discerned.

In many cases well-defined dark, linear markings, with lobate margins radiate from a central dark caldera (Plate 2a, p. 783). These are presumed to be dark flows because of their association with a caldera and the strong resemblance to the pattern of terrestrial lava flows. Examples are at 35° S–353° W, 85° S–340° W, 37° S–5° W, 15° S–322° W, and 52° S–330° W. In all these cases the flows are long and narrow and the array is markedly radial. There seems to be a succession of flows, the more recent flows being darker. The pattern is very similar to that on the flanks of large terrestrial shield volcanoes such as Mauna Loa (Hawaii) and Fernandina (Galapagos). One respect in which ionian flows appear distinctly different is that the terrain near parts of some flows is very bright. The brightenings commonly trace the edge of the flow and are clearly related to it. The general effect is as though the dark flows were backlit.

Other dark markings around the calderas may be very broad yet still show the lobate margins characteristic of flows. Examples are at 23° N–277° W and 11° N–272° W (Plate 1). These flows resemble the long narrow ones but are broader, perhaps because the slopes over which they have flowed are relatively shallow.

Several well defined bright, lobate markings which originate at a caldera are also observed. These are interpreted as light flows. One originates at a caldera at 7° S–286° W (Plate 2) and appears to have flowed south-west 250 km partly flooding two other calderas. Many other calderas are surrounded by bright radial markings with sharply defined edges. These are also probably volcanic flows. Examples are at 8° S–342° W and at 29° S–328° W.

In the most common pattern of markings around the caldera, dark and light markings, while still elongated radially, interweave so that individual units cannot be traced for any great distance (Plate 5, p. 786). Whether we are observing light flows on a dark background, dark flows on a light background, or both, is unclear, but the presence of both light and dark flows seem likely in many cases. Such intricate albedo patterns are very common in the region between 0° and 40° S and 280° W and 20° W. They are far less common south of 40° S at the same longitudes.

Another recurrent albedo pattern is one in which a sharply demarcated bright region surrounds a caldera, as though the surface were bleached. The bright region generally has a very sharp crenulated margin, distinctly different from the bright haloes that surround many features. A good example is at 2° N–2° W (Plate 2b). The origin of these markings is not known. Many calderas, therefore, are surrounded by albedo patterns that suggest the presence of both dark and bright flows and some are surrounded by a well defined bright region that apparently has no flow structure. In general, no relief can be seen on the albedo features.

Positive relief features: While most calderas do not seem to be within sharply defined edifices, a variety of positive relief features are recognised. Most are puzzling and difficult to relate to terrestrial landforms. Among the more comprehensible, because of their resemblance to low volcanic cones, are two pancake-like constructions at 23° S and 26° S at longitude 345° W (Plate 5, p. 786). They are nearly circular, approximately 150 km across, and surrounded by a low escarpment. Each has a bright-floored small crater in the middle. The albedo of the main edifice is uniform and close to that of the surroundings. The pancakes appear to transect and thus be younger than fields of flows from nearby calderas. Similar features occur at 1° N–316° W and at 3° N–326° W. Other positive features surrounded by a low escarpment and with small central craters differ in being markedly non-circular and have ridges on the surface that trend roughly radial to the central crater.

A second type of positive feature is typified by a butterfly shaped structure at 58° S–350° W (right-hand corner, Plate 3, p. 784). At the head is a caldera very similar to the other ionian calderas. The wings are formed by several broad sheets up to 300 km across, each surrounded by a low escarpment. The sheets partly overlap one another and seems to be superimposed

on a dark tail apparently comprised of flows. The sheets themselves have no distinctive albedo pattern and are very similar in brightness to the surrounding plains. Similar sheet-like features are associated with calderas at 18° S–350° W, 42° S–358° W, 10° S–295° W, and 32° S–330° W. Several tongue-like sheets, particularly near the south pole, resemble those around the butterfly but cannot be traced to any specific caldera.

Rugged mountains with high relief, best seen close to the terminator, constitute a third class of positive feature (Plate 3). The highest resolution view of one of these mountains at 68° S–35° W shows that the slopes of the mountain have a distinct texture formed of at least two intersecting sets of parallel lineaments. Accurate elevation measurements have yet to be made, but the relief seems large. This particular mountain has a caldera at the base of the south flank. Although the morphology of the mountain itself is not distinctively volcanic, a volcanic

Generally the diffuse halo is transected by dark deposits on the caldera floor but some caldera floors lack dark markings and instead have diffuse bright markings which appear to originate at the walls. Bright haloes are also associated with many of the dark flow-like features that radiate from some calderas (Plate 2). In some cases the bright markings provide a backdrop to the entire flow field, in other cases just one or two of the flows are highlighted. Small, diffuse brightenings also occur in places along the low escarpments that bound the sheet-like features associated with some calderas. As indicated above, bright haloes also occur around some mountainous features.

Diffuse dark features are less common. Several calderas have roughly circular diffuse dark markings that overlap both the wall and the surroundings. Others have roughly radial dark markings very similar to those at plume centres. Other dark markings are very dense and irregular in shape.

One striking aspect of volcanic features on Io is their extreme youth. Johnson *et al.*⁴ estimates that Io is being resurfaced from plume activity alone at an average rate of 5×10^{-3} to 3×10^{-1} cm yr⁻¹, yet albedo patterns are clearly visible over much of the surface. A deposit of a few millimetres is sufficient to mask most albedo differences except where there is significant micro-relief. The calculated resurfacing rates, therefore, imply that the albedo features are at most only a few tens of years old and may be only a few months old. Topographic features may survive longer but the lack of impact craters larger than 1 km implies that 100 m of relief is removed in $< 10^6$ yr (ref. 4). Relative ages can be inferred from transection relations between different flows and between flows and calderas. In general, albedo contrasts seem to decrease with increasing age as dark areas become brighter and bright areas darken. The impression is consistent with ageing mechanisms such as steady burial sulphur rich compounds (SO₂ frost, elemental sulphur) or tephra, or physical and chemical alteration of the rock forming materials.

Discussion

In the following we will examine what the high rates of volcanic activity and apparent abundance of sulphur on and around Io imply about the type of volcanism on Io. We assume that the snapshot view of Voyager is typical of much of Io's geological past, and that it has sustained a high level of volcanic activity for much of its history. The assumption is consistent with tidal dissipation providing the required energies¹⁷.

Are the flow-like features seen at the surface mostly silicate lavas or can they be explained solely by sulphur or sulphur-rich compounds? There is evidence to support both these hypotheses and to suggest that both silicate and sulphur rich flows may exist.

Sulphur lava could in principle be juvenile, that is, derived by separation of a sulphur-rich melt from a primitive magma. Such flows are however, unlikely. The solubility of sulphur in mafic magmas has been studied in detail¹⁷. The solubility depends mainly on the Fe content of the magma and the S₂ and O₂ fugacities. Depending on the value of these variables, 0.1–1% sulphur can dissolve in the silicate melt. At higher values of sulphur a sulphide melt separates out so that the silicate and sulphide melts coexist as immiscible liquids. Once this state is reached addition of more sulphur has little effect on the silicate phase relations, except insofar that Fe is partitioned between the sulphide and silicate liquids. The sulphide melt also acts as a sink for metals such as Cu, Ni, Co and Cr. Separation of sulphide globules from juvenile basalt melts was reported by Skinner and Peck^{19,20} in samples from the Alae lava lake, Hawaii, and sulphide globules have been analysed from deep sea basalt^{21–24}. Such sulphide melts, however, are unlikely to be the source of sulphur-rich lavas at the surface of Io. In a closed magma chamber, the sulphides tend to sink and form sulphide-rich layers such as are common at the base of many layered intrusions. During the early planet-wide melting postulated by Peale *et al.*¹⁷ such sulphide melts may have separated out to form a core. If the magma rises to the surface, however, it tends to

Table 1 Location of volcanic features*

Type	Voyager Picno	Lat	Long
Long, narrow dark flows	0073J1+000	35° S	353° W
	0147J1+000	85° S	340° W
	0184J1+000	37° S	5° W
	0043J1+000	15° S	322° W
	0199J1+000	52° S	330° W
Broad dark flows	0027J1+000	23° N	277° W
	0027J1+000	11° N	272° W
Light flows	0041J1+000	8° S	342° W
	0135J1+000	29° S	328° W
Bright calderas	0073J1+000	28° S	338° W
	0073J1+000	62° S	20° W
Caldera with concentric markings	0027J1+000	23° N	277° W
Calderas with albedo of surroundings	0077J1+000	74° S	340° W
	0077J1+000	50° S	12° W
Calderas with sheet flows	0073J1+000	18° S	350° W
	0073J1+000	42° S	358° W
	0051J1+000	10° S	295° W
	0137J1+000	32° S	330° W
	0093J1+000	1° N	316° W
Pancake-like constructions	0131J1+000	3° N	326° W
	0119J1+000	43° S	265° W
Tectonic uplifts	0109J1+000	52° S	280° W

* Dark calderas are too numerous to list. For location of active plumes see ref. 10. Errors in location may be as large as 15°.

origin is suggested by a bright halo that completely surrounds the mountainous area. Not all the mountains have bright haloes. Some are surrounded by sheet-like deposits similar to those described in the previous paragraph.

In a fourth type of positive feature, an originally smooth and level surface appears to have been upbowed and cracked along radial faults (Plate 4, p. 785). A smooth surface, broken only by the radial graben, slopes gently away from the centre which is generally occupied by an irregular depression connected to the radial graben. The largest is at the centre of the heart-shaped structure at 30° S–250° W, and is the source of the largest plume on Io¹. Others are at 43° S–265° W and 52° S–280° W.

Diffuse markings: Diffuse dark and light markings are common in all the regions so far examined and are generally associated with some topographic feature. Near-vertical views of active plumes commonly show a radial dark and light pattern at the plume centre which merges outwards into a bright halo, which in some cases shows a concentric banding¹⁶. Although the active plume itself must contribute to the appearance from above, most of the diffuse markings are probably on the surface. Many similar markings not now associated with plumes may have resulted from previous activity.

Many calderas have diffuse bright markings. Most commonly the bright patterns around the caldera are asymmetric although some, such as those at 66° S–10° W and 74° S–0° W, are surrounded by a fairly uniform halo. In some places dark streaks, similar to those at plume centres, extend away from the caldera.

degas and lose much of its sulphur, which condenses on the surface or in voids within the near surface rocks. Most terrestrial basalts thus lack sulphide globules. Exceptions are those erupted on the ocean floor, where degassing is prevented by the high confining pressure.

Accumulation of sulphur at the surface of Io probably resulted from pervasive outgassing of the interior. The results would be similar whether the outgassing occurred early as a result of runaway melting of the interior as suggested by Peale *et al.*¹⁷ or whether the outgassing occurred more slowly as a result of progressive melting. Many of the lighter volatiles such as water would be lost to space whereas the heavier volatiles, such as sulphur would mostly condense in the near surface regions⁴. The net result would be interbedded lavas and sulphur-rich deposits near the surface. If subsequent volcanism was of sulphur alone then the sulphur would ultimately segregate into a near surface layer. If, however, both silicate and sulphur volcanism continued, as we suggest, then the near surface stack of interbedded lavas and sulphur-rich deposits would be maintained. Near surface volcanism would result in progressive burial of the interbedded sequence until temperatures are reached such that the sulphur melts or volatilises. Peale *et al.*¹⁷ estimate that silicate melting temperatures are reached at depths of 18 km. This implies that sulphur is above its melting temperature below depths of 3–4 km if the crust is primarily silicate. In the upper part of the crust therefore, sulphur would be relatively mobile, being continually transported upwards from the base of the interbedded sequence to be redeposited in the cooler upper parts or on the surface. The silicates may also be recycled. Continued eruption of silicates onto the surface or into the near surface rocks could result in burial of the silicates to depths where they too would be melted and remobilised. The silicates may thus also be recycled, although on a time scale considerably longer than the sulphur. Such a model is consistent with many of the volcanic features seen on the surface.

Many of the flows on Io's surface may be composed of sulphur. Eruptions of molten sulphur have been described from the Siretoko-Iosan volcano in Japan²⁵ and a small sulphur flow occurs on the flanks of Mauna Loa, Hawaii²⁶. During the 1936 Siretoko eruptions, chocolate brown sulphur erupted from a vent on the side of the volcano and fed a brown flow, 20–25 m wide, 1.4 km long, and up to 5 m thick. The flow travelled down a valley that was filled with previous sulphur flows to a depth of several tens of metres and sulphur sublimates cover a considerable area around the vent and adjacent to the flow. The Hawaiian flow is only 5 by 10 m across and up to 45 cm thick, but it may have been partly covered by a younger flow²⁶. Both the Siretoko and Mauna Loa sulphur flows have been ascribed to remobilisation of fumarolic sulphur as ground temperatures in the fumarolic areas were raised above the sulphur melting point.

In view of the apparent sulphur-rich environment on Io, and extensive evidence of fumarolic activity, an analogous origin for flows on Io is reasonable. Some of the flows are over 300 km long. Whether sulphur could continue to flow over such distances would depend partly on whether a crust of low thermal conductivity forms over the flowing surface and whether lava tubes form as with basaltic lava flows. Although Watanabe²⁵ described various surface textures such as pahoehoe and aa on the sulphur, he does not discuss the properties of the lava crust. Flowing basalt generally develops a crust which is extremely vesicular as a result of outgassing of the magma during cooling. The crust generally has a very low thermal conductivity which retards loss of heat from the flow. In many cases the flows crust over completely and the lava flows on a tube below the surface. As a result flow can be sustained over great distances. It is not known whether a sulphur flow in Io would vesiculate in a similar manner. Flow over long distances is, however, aided by two other factors. First, the temperature of sulphur flows is very low compared to the silicate melts. Since heat loss varies as the fourth power of the absolute temperature the rate of loss of heat from a sulphur flow would be two orders of magnitude less than from a silicate flow, all other factors being equal. Second, the

peculiar viscous properties of sulphur may help sustain flow, at 200 °C the viscosity of liquid sulphur is 2.15×10^4 cP. On cooling the viscosity decreases such that at 150 °C the viscosity is only 7 cP. If the flows are sulphur then the differences in albedo between the flows may result simply from the different sulphur allotropes⁷. Although the sulphur flows at Siretoko revert rapidly to the yellow monoclinic form, sulphur flows on Io, because of rapid freezing to 130 K, could exist metastably in other states with a variety of colours and albedos⁷.

An alternative explanation of the flows is to attribute them to normal lavas, probably basaltic in composition. The variations in albedo would then result from differences in the amounts of volcanic sublimates on the surface or from differential chemical alteration of the flow surfaces. Such deposits or alterations are not uncommon on terrestrial lava flows. The flow itself, including its surface texture, is little affected but its reflectivity and colour may be dramatically affected by alteration and deposition.

The relief of the ionian surface supports the hypothesis that some silicate volcanism has occurred. Several rugged mountains with relief in excess of 10 km are probably composed of silicates. More importantly, however, with respect to the type of volcanism, is the relief within the calderas. Some have steep inward facing scarps that appear to be as much as 2 km high. This implies a significant yield strength⁷. While silicates certainly have the necessary yield strengths, it is not known whether sulphur could support cliffs of such magnitude at 130 K. If not, then the amounts of silicates erupted must be at least comparable to the amounts of sulphur.

The diffuse light markings are less controversial. They are probably sublimates, principally sulphur of SO₂, and could result from either silicate or sulphur volcanism. The origin of the diffuse dark markings is less clear. One possibility is that they are formed by volcanic ash. The ash could be produced simply by erosion of the vents by gas jets that form the observed plumes. Alternatively, the ash may form by rapid degassing of near surface silicate lavas charged with S or SO₂. The anhydrous nature of Io's surface^{27,28} suggests water is not involved. Unless the silicate lavas were juvenile such degassing would require that lava assimilate sulphur at intermediate depths and discharge the gas near the surface as both the pressure and the sulphur fugacity fall.

From these rather qualitative arguments we conclude that while there are strong arguments for sulphur flows on the surface of Io, silicate volcanism of the type familiar to us on Earth may also occur. The possibility of both silicate and sulphur volcanism raises some questions concerning the loss of heat from the planet. Repeated eruptions of large amounts of sulphur onto the surface would have an effect similar to increasing the thermal conductivity of the rocks which lie above the zone of sulphur melting. If tidal energy is dissipated deep within the body, that is the heat sources are all below the surface layer not within it, then the effect of increasing the thermal conductivity is to decrease the near surface temperature gradient and so increase the depth to melting of both sulphur and silicates. Large amounts of near surface sulphur would therefore tend to suppress silicate volcanism by forcing silicate magma sources to greater depths. The larger the amount of sulphur transported to the surface the greater the effect.

Johnson *et al.*⁴ estimate that Io is being resurfaced from plumes at a rate of 5×10^{-3} to 3×10^{-1} cm yr⁻¹. Assuming the plumes are due entirely to sulphur vaporised at 717 K then brought to the surface and condensed at 130 K, then the rate of loss of heat by vaporisation and cooling of sulphur alone is $1.7\text{--}100$ erg cm⁻² s⁻¹ at these resurfacing rates. This compares with 500 erg cm⁻² s⁻¹ required to remove the tidal dissipation energy given to Peal *et al.*¹⁷. Plume activity even at the upper limit of the rates given by Johnson *et al.*⁴ thus leaves ample energy to drive silicate volcanism.

The lack of large volcanic edifices may result both from isostasy and the lack of pressure to pump magma to a significant elevation above the general ionian surface. High rates of volcanic

activity imply melting at relatively shallow depths and the large size of many of the ionian calderas suggests magma reservoirs with lateral dimensions greater than their depth of burial. Both these factors should result in efficient isostatic compensation. This alone may explain the lack of large volcanic edifices although the presence of several rugged mountains is puzzling. Another factor may be the lack of density contrast between the lava and crustal rocks. Magma is pumped to the top of large

terrestrial shield volcanoes partly by the density contrast between the mantle and basaltic lava. On Io, however, the high rates of volcanism imply that most lava is derived by the remelting of formerly erupted materials. The composition of the lava is therefore probably similar to the surrounding rocks and the density contrast small. This, coupled with the relatively shallow depth of origin of the lava, may prevent generation of the pressure head needed to build large edifices.

Received 16 July; accepted 31 July 1979.

1. Morabito, L. A. *et al. Science* **204**, 972 (1979).
2. Smith, B. A. *et al. Science* **204**, 951–972 (1979).
3. Hanel, R. *et al. Science* **204**, 972–976 (1979).
4. Johnson, T. V., Sagan, C. & Soderblom, L. A. *Nature* **280**, 746–750 (1979).
5. Kupo, I., Mekler, Y. & Eviatar, A. *Astrophys. J. Lett.* **205**, L51–L53 (1976).
6. Broadjoot, A. L. *et al. Science* **204**, 979–982 (1979).
7. Sagan, C. *Nature* **280**, 750–753 (1979).
8. Smith, B. A. *et al. Nature* **280**, 738–743 (1979).
9. Fanale, F. P. *et al. Nature* **280**, 761–763 (1979).
10. Nash, D. B. & Nelson, R. M. *Nature* **280**, 763–766 (1979).
11. Smythe, W. D., Nelson, R. M. & Nash, D. B. *Nature* **280**, 766 (1979).
12. Hapke, B. *Geophys. Res. Lett.* (in the press).
13. Pearl, J. C. *et al. Nature* **280**, 755–758 (1979).

14. Williams, H. *Bull. Univ. Calif. Dept. Geol. Sci.* **25**, 239–346 (1941).
15. Pike, R. J. *Proc. 9th Lunar planet. Sci. Conf.* 3239–3273 (1978).
16. Strom, R. G. *et al. Nature* **280**, 733–736 (1979).
17. Peale, S. J., Cassen, P. & Reynolds, R. T. *Science* **203**, 892–894 (1979).
18. Haughton, D. R., Roeder, P. L. & Skinner, B. J. *Econ. Geol.* **69**, 451–467 (1974).
19. Skinner, B. J. & Peck, D. L. *Econ. Geol.* **61**, 802 (1966).
20. Skinner, B. J. & Peck, D. L. *Econ. Geol.* **4**, 310–322 (1969).
21. Moore, J. G. & Calk, L. C. *Am. Miner.* **56**, 476–488 (1971).
22. Moore, J. B. & Favvi, B. P. *Contr. Miner. Petrol.* **33**, 118–127 (1971).
23. Matthey, E. A. *J. geophys. Res.* **81**, 4269–4275 (1976).
24. Czamaske, G. J. & Moore, J. G. *Bull. geol. Soc. Am.* **88**, 587–599 (1977).
25. Watanabe, T. *Jap. J. Geol. Geogr.* **17**, 289–310 (1940).
26. Skinner, B. J. *Pacif. Sci.* **24**, 144–145 (1970).
27. Fink, V., Dekkers, N. H. & Larson, H. P. *Astrophys. J. Lett.* **179**, L155–L159 (1973).
28. Cruikshank, D. P., Jones, T. J. & Pilcher, C. B. *Astrophys. J.*, **255**, L89–L92 (1978).

Volcanic eruption plumes on Io

Robert G. Strom*, Richard J. Terrile†, Harold Masursky‡ & Candice Hansen†

*Department of Planetary Sciences, University of Arizona, Tucson, Arizona 85721

†Jet Propulsion Laboratory, Pasadena, California 91103

‡US Geological Survey, Flagstaff, Arizona 86001

An umbrella-shaped plume detected above Io confirms that Io is volcanically active. Preliminary analyses of eight such eruptive plumes are presented.

THE detection of an umbrella-shaped plume extending about 280 km above the bright limb of Io¹ was one of the most important discoveries made during the Voyager 1 encounter with the jovian system. This discovery proves that Io is volcanically active at present, and the number and magnitude of these eruptions indicate that Io is the most volcanically active body so far discovered in the Solar System. The source of the heat which drives this volcanic activity is probably the tidal heating mechanism proposed by Peale *et al.*².

The Voyager 1 Io data reveal eight eruptive plumes detected on images from 4 days before to 3 days after the Io encounter. Up to 4 days before encounter the resolution was too poor to detect the plumes. In general, the eruptive plumes can only be seen near the bright limb or at small phase angles near the dark limb where they are projected against the dark sky and shine by reflected sunlight. However, most of the plumes are so faint that special processing is required to detect them against the black sky.

The Voyager 1 longitude coverage at the limb is limited. Depending on resolution, at 10° in front of or beyond the bright limb, about 10 km of a 40-km high plume would be visible against the black sky. The surface area between 10° either side of the limb covered by the Voyager 1 pictures with resolution better than 90 km per television line pair represents about 80% of the surface of Io. About 54% of the surface was seen in the limb regions at 10 km per television line pair resolution or better. Considering the faintness of the plumes we feel that only about half of the planet has been surveyed satisfactorily, and other eruptive plumes, particularly smaller ones, may have passed undetected.

Characteristics of plumes and their sources

Figure 1 shows the distribution and approximate diameters of the eight detected plumes, while Table 1 lists their dimensions,

locations, and the times at which they were seen. The exact location of Plume 8 is uncertain because of limited photographic coverage but it is located along the error bar in Fig. 1. The dimensions of Plumes 5, 7 and 8 are not well established because of poor resolution and/or the faintness of the plume. Except for the faint Plume 8 at latitude -45° , all observed plumes occur within $\pm 30^\circ$ of the equator.

Plume 1, the largest, is about 1,000 km wide and 280 km high measured in both clear and UV filters. It is a highly symmetrical umbrella-shaped cloud with a central fountain about 35 km wide near its base (Plate 5b, p. 786). When viewed nearly vertically (Plate 6a, p. 788) the plume source is surrounded by an elliptical, heart-shaped diffuse area. The centre of this area contains one or two black spots from which the eruption apparently originates. Dark diffuse streaks radiate from these black spots for a distance of about 150 km. A light zone in which topographic features can be seen more clearly than in other areas surrounds the dark streaks. This light zone is surrounded by a dark diffuse region whose outer limit approximately coincides with the plume width. A light diffuse band about 90 km wide occurs on the western edge of the dark region, while on the north and east the dark zone grades outwards into a very diffuse region about 300 km wide. The dark and light diffuse zones which comprise this heart-shaped region probably represent airborne and/or surface pyroclastics. Plate 6b, p. 788 is a high resolution (1.5 km per television line pair) oblique image of the source region of Plume 1. In this view the sky was processed separately to show the filamentary structure in the plume off the limb of Io. The dark diffuse radial streaks which form the centre of the plume are associated with three black spots located at the northern end of a highly fractured uplift about 200 km wide. The central portion of this uplift is bisected by a graben 25–50 km wide. The largest black spot is adjacent to a basal scarp encircling the uplift and appears to be associated with, or form part of, an adjacent caldera. Two smaller black spots occur along the basal scarp and a parallel fracture, and appear to be the sources of the dark diffuse material to the west. It is not clear which of these three vents is the source of Plume 1; possibly all three are erupting simultaneously. Several smaller dark diffuse

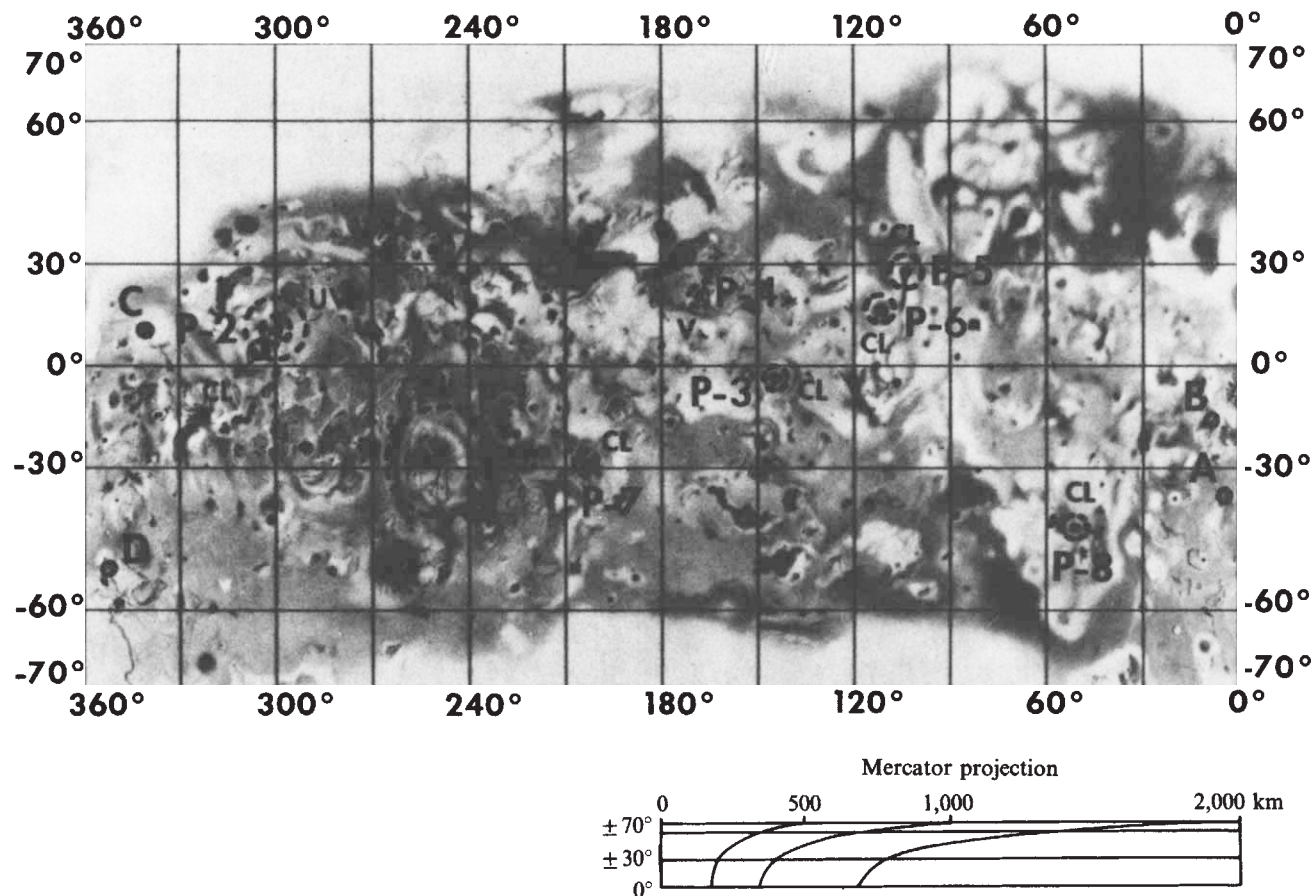


Fig. 1 The location of the eight eruptive plumes and their diameters. Four possibly active eruption sites are indicated at A–D. The latitude/longitude grid is very approximate and contains errors of 10° or more. ●, Eruption plume source; ○, plume diameter; CL, clear filter, UV, UV filter; V, violet filter; O, possible eruption.

areas, some with bright spots as their source, are probably eruptive centres which may have been active when the picture was taken. The large scale symmetry of the plume suggests that the geometry of the vents is very simple, suggestive of pipe vents or short fissures. All are located along fractures or the basal scarp which itself is probably a fault scarp.

The character of Plume 2 (Plate 5b) is different from that of the other plumes. It is a very diffuse cloud lacking the umbrella shape or columnar fountain characteristic of most other plumes. Viewed through the clear filter it is about 100 km high and 210 km wide, but in UV light it is more than twice as large (Fig. 1). Therefore, this plume shows two components: a central core seen in visible light surrounded by a much fainter envelope of material which scatters preferentially in the UV. The central core is probably dark pyroclastic ejecta while the faint envelope may be a reflection from extremely fine particles or possibly resonance scattering by gas³. Other plumes such as Plumes 1 and 4 were also observed in the UV but did not show evidence of larger envelopes in this filter. Plate 5a, p. 786 is a moderately high resolution picture (~3 km per television line pair) of the source region of Plume 2. The eruption is associated with a black linear strip (A in Plate 5a) about 175 km long and 25 km wide. Dark jets of material, probably pyroclastics, appear to be erupting from the west end of the black strip. These streaks form a dark patch with a maximum dimension of about 200 km. This coincides approximately with the width of the core of the plume as viewed in visible light. Although the west end of the black strip seems to be the main eruption site, smaller dark and light streaks issuing from the northern edge of the strip strongly suggest that eruptions occur along its entire length. A bright zone about 420 km in diameter surrounds the main eruption site. Another brightish region is associated with the northeastern part of the black strip. Surrounding these bright zones is a very diffuse region about 800 km in diameter. The outer edge of

the plume viewed in UV light lies about halfway between the bright zone and the outer limits of the darker diffuse region. Plume 2 appears to be the result of a fissure eruption occurring along a fracture 175–200 km long. This very long fissure eruption may account, at least in part, for the lack of a columnar fountain and the very diffuse appearance of the plume viewed in profile. The dark material forming the black strip has an irregular outline suggesting it is made up of lava flows which have issued from the fissure.

Plume 3 is a very symmetrical umbrella-shaped plume about 250 km wide and 70 km high (Plates 5b, 6a and 7a, pp. 786–789). It is associated with a dark circular area surrounded by a bright white ring shown in Plate 7b, p. 789. The brightest part of the central fountain is about 15 km in diameter and coincides with a dark spot at the centre of the inner ring (arrow in Plate 7a). The diameter of the plume is about the same as that of the inner dark ring, suggesting that most of the deposition from the plume is concentrated within this area. Plate 7a is an oblique view of the eruption showing jets of material being ejected on ballistic trajectories. Some of this ejecta clearly lands on the white ring. Whether the bright ring surrounding the source region is associated with this eruption or is an earlier deposit is not clear. Johnson *et al.*⁴ speculate that bright rings may be SO₂ frost deposited from the volcanic gas. The narrowness of the central fountain (~15 km) and the almost perfect symmetry of the plume suggest that this eruption is either a short fissure or more probably a pipe vent eruption.

Plumes 4–8 are not as well documented as Plumes 1–3. Plume 4 (Plate 5b) is a high (~95 km), relatively thin column capped by a diffuse asymmetrical cloud elongated to the north. This plume has the largest height-to-width ratio (1.2) and is associated with the southern end of a dark band where dark radial streaks are surrounded by a lighter diffuse region (Plates 6a and 7b). Plume 5 was photographed only at very low resolution or when it was

Table 1 Voyager 1 plume observations

Plume no.	Location* lat/long	No. of observations†	Time of first observation	Time of last observation	Duration	Plume* width (km)	Plume* height (km)	Central fountain* width (km)	Plume height/ plume width
1	-29°/249°	4	-16.9h	+3d 1.4h	3d 18.3h	10000(CL)	280(CL)	35(UV)	0.3
2	+10°/300°	8	-3d 6.0h	+3d 1.4h	6d 7.4h	210(CL) 585(UV)	100(CL) 210(UV)	—	0.5(CL) 0.3(UV)
3	-4°/144°	4	-1d 3.8h	-9.3h	18.5h	250(CL)	70(CL)	15(CL)	0.3
4	+21°/169°	2	-1d 1.6h	-5.6h	20.0h	75(V)	95(V)	20(V)	1.2
5	+27°/105°	4	-1d 11.3h	-11.1h	1d 0.2h	200(CL)	80(CL)	?	0.4
6	+17°/111°	3	-1d 9.9h	-11.1h	22.8h	250(CL)	80(CL)	60(CL)	0.3
7	-29°/203°	2	0.0h	+0.5h	0.5h	180(CL)	120(CL)	?	0.6
8	-44°/ 52°	3	0.0h	+2.0h	2.0h	150(CL)	70(CL)	15(CL)	0.4

CL, clear filter measurement; UV, ultraviolet filter measurement; V, violet filter measurement.

* These values are very approximate pending additional data reduction and more accurate measurements.

† Each observation may be one image or a sequence of several images taken through different filters.

barely protruding above the limb. It has about the same dimensions as Plume 6 and is located about 300 km to the north-east. A dark diffuse region surrounded by a lighter halo is the source area of this plume (*Plate 7b*). Plume 6 is unusual in that it has a very broad central fountain about 60 km wide and 80 km high, surrounded by a diffuse cloud about 250 km wide (*Plate 5b and 7a*). This plume is located over a broad linear black band surrounded by a dark diffuse region about 200 km in diameter (*Plate 7b*). This dark region is ringed by a light area which in turn is surrounded by a very diffuse circular region with an albedo similar to that of the inner ring. The width of the central fountain is similar to the projected (line-of-sight) length of the central black band suggesting that Plume 6 is the result of a fissure eruption fountaining along the length of the black band. The width of the diffuse cloud surrounding this fountain is approximately the same as the outer diameter of the bright ring. Plume 7 (*Plate 5b*) is a tall (~120 km) relatively thin column with a diffuse top about 180 km wide. It has the second largest height-to-width ratio and is associated with a three-component, black diffuse region best seen in *Plates 6a and 7a*. The northern and southern components are circled by a bright ring which in turn is surrounded by a darker diffuse region about 400 km in diameter. This plume has not yet been located accurately enough to determine which of the three components is the eruptive vent. However, both the northern and southern components seem to have dark jet-like streaks and surrounding bright rings suggesting they are the most likely sources. Plume 8 (*Plate 5b*) is very faint and was repeatedly viewed on images with nearly the same viewing geometry. The difficulty in triangulating from these images and the poor resolution coverage of the source region make the location of this plume uncertain (*Fig. 1*).

Other possible eruptions

All plume sources are characterised by: (1) a very dark or black central area consisting of radial to sub-radial jet-like streaks; (2) an irregular to circular bright ring surrounding the central core; and (3) a very broad diffuse region surrounding the bright ring. Numerous regions, other than plume sources, display these same characteristics. Several of these areas are indicated by arrows in *Plate 7b*. Some of these regions, comparable in size to plume sources, were photographed at the limb but show no plumes. Most were not viewed at the limb and may or may not have been active during the Voyager 1 encounter. Four possibly active eruption sites photographed at high resolution are indicated in *Fig. 1*. All are characterised by dark jet-like streaks very similar to those associated with Plume 2 and photographed at comparable resolution or better. None of these areas was seen at the limb. Furthermore, numerous small dark or bright diffuse spots and streaks seen on high resolution pictures (better than 4 km per television line pair) may have been active during the encounter. Several of these are shown in *Plate 6b*. The presence of eight enormous eruptions occurring simultaneously almost certainly implies that small eruptions below the detection limits of the television system were also taking place.

There is evidence that at least one region on Io changed substantially during the Voyager 1 flyby. The appearance of the region surrounding Plume 4, particularly to the north, has changed dramatically during the interval between the shuttering of the pictures shown in *Plate 7a and b*. In *Plate 7b* numerous black spots and bands to the right of A are no longer visible in *Plate 7a* taken 5.5 h later. These changes may be due to obscuration of surface detail by airborne or surface eruption products, or they may be a phase effect due to a difference in the scattering properties of the material in this region. *Plate 7a* was taken at a phase angle of 2.5° while *Plate 7b* was taken at a phase angle of 14.8°. These changes are unlikely to be solely the result of the eruption of Plume 4 because of its small size and great distance from the altered areas (see *Fig. 1*). If this obscuration is caused by an eruption then its source is probably at or near the hub of the dark radiating bands. Furthermore, it must have been either initiated or intensified some time during the 5.5-h period between *Plate 7a and b*. Plume 4 was last seen about 5 h after *Plate 7a* was taken. An eruption large enough to cause these changes should have produced a plume large enough to be visible on that picture. The absence of such a plume suggests that either a large eruption began and terminated within a 10-h period or that the changes are indeed the result of a phase effect. Comparative photometric studies of plumes and this region may help decide between these alternatives.

Duration of eruptions

Eruption plumes were observed by Voyager 1 for 6.5 d; the time during which the resolution was sufficient to detect them. However, not all plumes were observed for this period of time because of limited limb coverage and poor resolution. Table 1 shows the relative sequence of plume observations. Only Plume 2 was observed periodically over this entire period. In all instances where plumes were observed on the limb over several days no detectable changes were seen in the size or intensity of the eruption. However, observations of Plume 3 against Io's disk do suggest a change in the eruption rate between the vertical view in *Plate 7b* and the oblique view in *Plate 7a*. Plume 3 in *Plate 7a* appears darker and more optically thick than in *Plate 7b*. If these changes are real and not an effect of the changing viewing geometry then this implies a large change in the eruption rate over a period of 5.5 h. Plume 3 was observed erupting on the limb about 15 h before *Plate 7b* was acquired and about 2 h after *Plate 7a*.

In an attempt to place limits on the extent and duration of this volcanic activity Voyager 2 will observe Io intermittently for 4 d at a range of $4.76\text{--}1.38 \times 10^6$ km, with a corresponding resolution of 88–26 km per television line pair. Io will be imaged continuously for 6.5 h at a range of 1.13 to 1.19×10^6 km with 21 km per television line pair average resolution. Plumes 5 and 6 will cross the bright limb during this period.

Age of recent eruptions

Although most of the numerous circular diffuse areas with black centres may not have been active during the Voyager encounter, they almost certainly represent the former sites of relatively

recent eruptions. In some cases their relative age can be determined by superimposing one diffuse ejecta deposit on another. For example, in *Plate 7b* the ejecta deposit of the eruption site just north of Plume 3 overlies that of the eruption site to the north, indicating that it is younger.

An estimate of the upper limits of the absolute age of relatively recent eruptions can be determined by the volcanic resurfacing rate estimated by Johnson *et al.*⁴. From the lack of impact craters and from calculations for the observed volcanic eruptions they estimate a global deposition rate of about 1 mm yr⁻¹ or more. Of course, the local deposition rate will be much greater or less than this value, depending on whether the area is close to or remote from an eruption site. Based on the relatively limited limb coverage, the observed eruption plumes seem to be concentrated within $\pm 30^\circ$ of the equator. Also, the eruptions appear to be more or less randomly distributed with respect to longitude (Fig. 1). This distribution may be linked in some complex way to body tides on Io. If the equatorial concentration of eruptive plumes is real, then either the intensity or frequency of eruptions, or both, are greater in the equatorial regions than in the polar regions. This would imply that the average deposition rate in the equatorial areas is greater than that in the polar regions. The dark polar regions and light equatorial areas may be partly a result of this difference in deposition rate; the lighter equatorial areas resulting from a greater deposition rate of light pyroclastics and sulphur compounds. This suggests that on average the surface in the equatorial regions is younger than that in the polar regions.

A more or less uniformly distributed layer of pyroclastics ≤ 1 cm thick should be sufficient to make the boundary between the diffuse ejecta deposits associated with eruption sites and the surrounding terrain indistinguishable. Therefore, if the depositional rate estimated by Johnson *et al.*⁴ is applicable to at least the equatorial regions, then on average the eruption sites with well-defined boundaries, such as those indicated in *Plate 7b*, must be extremely young (possibly ≤ 10 yr). The large areal dimensions of these deposits and their variation in colour and albedo suggest that large-scale changes in colour and albedo patterns take place over relatively short periods of time; perhaps a few years or shorter. Integrated full-disk variations of up to 10% are consistent with the historical data⁵.

Conclusions

The largest of the eight volcanic eruption plumes (Plume 1) is about 280 km high and 1,000 km in diameter. When viewed nearly vertically, the eruption sites are characterised by a central

dark or black irregular to circular spot usually with radiating dark jet-like streaks. This is surrounded by a series of broad light and dark diffuse rings. The morphology of the plumes seems to be related to the dimensions and geometry of the eruption vent. Very diffuse plumes with no discernible central fountain (Plume 2) or a very broad one (Plume 6) are probably associated with long fissure eruptions. Highly symmetrical, umbrella-shaped plumes with relatively narrow central fountains (Plumes 1 and 3) may be related to either short fissure or pipe vent eruptions.

Many areas on Io display the same surface characteristics as those associated with plumes, suggesting they are also active or former eruption sites. Some of these were viewed at the limb but showed no plumes; others were not photographed at the limb and may have been active during the encounter. Most of the larger of these eruption sites seem to occur in the equatorial regions. Even if these sites were not active during the Voyager 1 encounter, they are all probably the locations of relatively recent eruptions. The pyroclastic deposition rate estimated by Johnson *et al.*⁴ suggests that they may be, on average, 10 yr old or less.

The appearance of the region north of Plume 4 changed significantly during a 5.5-h period. These changes may be due to a phase effect caused by different scattering properties of the surface in this region, or to obscuration of surface detail by eruption products from an eruption which took place during this period. If these changes are due to an eruption then the time coverage suggests that it was initiated or intensified, and then terminated or abated during a 10-h interval. This, with the long duration of other plumes, suggests that individual eruptions may be both short-lived and long-lived.

All of the observed plumes except Plume 8 are located in the equatorial regions between $\pm 30^\circ$ latitude. If this distribution is real then the intensities and/or frequency of eruptions is greater in the equatorial than in the polar regions. It follows that the equatorial surfaces are generally younger than those in the polar regions which in part may account for the difference in albedo between these two regions.

We thank the staff of the Image Processing Laboratory at Jet Propulsion Laboratory, especially Charles C. Avis, for assistance.

Received 16 July; accepted 20 July 1979.

1. Morabito, L. A., Synnott, S. P., Kupferman, P. N. & Collins, S. A. *Science* **204**, 972-976 (1979).
2. Peale, S. J., Cassen, P. & Reynolds, R. T. *Science* **203**, 892-894 (1979).
3. Smith, B. A. *et al. Science* **204**, 951-971 (1979).
4. Johnson, T. V. *et al. Nature* **280**, 746-750 (1979).
5. Morrison, D. *et al. Nature* **280**, 753-755 (1979).

Erosional scarps on Io

J. F. McCauley*, B. A. Smith† & L. A. Soderblom*

*US Geological Survey, Flagstaff, Arizona 8600

†Department of Planetary Sciences, University of Arizona, Tucson, Arizona 85721

Irregular or fretted scarps on Io are similar to those found on Earth and Mars. A sapping mechanism involving liquid SO₂ is proposed to explain these complexly eroded terrains on Io.

THE surface of Io exhibits a complex planetwide array of scarps that are typically hundreds of kilometres in length and estimated to be hundreds of metres high. At least four types of scarps can be recognised: (1) circular, semicircular to rectilinear scarps that encompass volcanic collapse depressions or calderas; (2) arcuate to digitate or locally bulbous scarps along flow fronts that

emanate from calderas and other vents; (3) long, straight to curvilinear tectonic scarps and narrow linear valleys that are interpreted as normal faults and graben; (4) irregular or fretted erosional scarps in layered terrains. The first three types of scarps are described in accompanying papers and will only be discussed here incidentally to the erosional scarps. Io and Mars are the only bodies other than the Earth which show complexly eroded scarps.

The present consensus about the fretted terrain on Mars is that it formed by internal sapping processes related to the melting of ground ice and the release of water to the surface^{1,2}. The water-poor, sulphur-rich surface of Io as revealed by Voy-

ager 1³ must have evolved very differently than the surface of Mars. However, there are many similarities in the erosional patterns observed. Mechanisms that require external activity such as fluvial and eolian processes are not considered likely for Io because of the exceedingly tenuous atmosphere⁴ and low surface temperatures. The extreme youth of Io's surface⁵ suggests that sputtering by particles in the jovian magnetosphere is also inadequate to explain the erosional features observed.

Description of scarps and erosional patterns

The most conspicuous stacks of layered terrain and associated erosional scarps lie in the south polar region of Io (see *Plate 3*, p. 784). Caldera scarps, flow-front scarps and tectonic scarps are also present. The flow-front scarps and tectonic scarps both bound flat-topped plateaus, the largest of which extend to ~10⁵ km². The surfaces of most of the plateaus appear to be similar to those of the surrounding lowland plains. Interspersed with the layered terrain and locally forming at its expense, are circular-to-irregular calderas, some of which are surrounded by either flat-topped or hummocky scarp-bounded flows. Thus, some combination of surface flow and tectonic processes is apparently required to explain these stacked deposits.

Our initial speculation was that the regional transfer of volatiles, probably from Io's eruptive plumes^{6,7}, condensation and subsequent sublimation were involved in the formation and subsequent erosion of the polar plains. The Voyager 1 infrared spectrometer (IRIS) discovered a tenuous SO₂ atmosphere near a plume in the equatorial region^{4,8}, and the observed partial pressure of the SO₂ (~0.1 μ bar) is thought to be in equilibrium with SO₂ frost on the surface in the equatorial region. Furthermore, SO₂ frost has been suggested as the cause of an absorption feature near 4.1 μ m seen in ground-based spectra^{9,10}. The equilibrium temperature at the poles is colder by many tens of degrees than at the equator, and with time the poles should act as cold traps. Some of the SO₂ gas produced in the eruptive plumes at lower latitudes should eventually migrate to and condense out in the poles. Thus, the polar plains and the higher scarp-bounded plateaus are probably abnormally rich in SO₂ just as the polar regions on Mars are rich in CO₂. However, sublimation of SO₂ to explain the erosion pattern in the polar regions of Io is not consistent with the geological relationships observed.

Transection and superposition relationships between individual scarps indicate a complex history of deposition, faulting and erosion. Complex cliff retreat patterns can be seen along the present edges of what are inferred to have been both tectonic and flow-front scarps (*Plate 3*). Extensive stripping back of the original surfaces is observed within the stacked layers. The scarp that surrounds this unit shows scalloped patterns consisting partly of intersecting semicircular alcoves. Somewhat rounded prongs stand out from the erosional fronts, and wedge-shaped reentrants contribute locally to scarp irregularity. Long gash-like to sometimes slightly-sinuuous valleys also extend into the edges of the layered terrain. Scattered, highly-irregular to near-equidimensional erosional remnants that are now completely separated from the main erosion front are also present. They consist in part of ragged, generally elongate mesas with flat tops which are concordant with the upper surface of the layer undergoing erosion. Even more eroded, elongate to equidimensional rounded hills of lower relief are scattered within the field of mesas near the centre of *Plate 3*. These mesas and hills are analogous to terrestrial inselbergs in their relation to the main scarp and are evidence of some type of pediplanation process on Io. The vague remnants of similar but more subdued scarps from beneath younger layers attest to multiple cycles of deposition and erosion.

Layered and eroded or fretted terrain is not restricted to the south polar region of Io. Smaller patches of flat-surfaced, layered terrain surrounded by lower but similarly eroded scarps are present in the mid-latitude and equatorial regions. The apparent lower relief of these scarps, however, may be a product of poorer illumination and viewing geometry of these pictures. Well-developed tracts of layered and eroded terrain occur near

10° S, 260° W under somewhat more favourable illumination, but outliers or inselbergs are absent, suggesting that the intensity of erosion falls off away from the poles. Where volcanic vents are tightly clustered and digitate flows of the type described by Carr¹¹ are abundant, layered and eroded terrain is not particularly evident.

The upper centre of *Plate 3* (near 50° S, 340° W) shows a complex stack of layered units associated with a dark-floored caldera. The uppermost layers show bulbous flow-front morphology and their tops are marked by orientated grooves and ridges, indicative of the flow of relatively fresh viscous materials. The lower units have more planar tops and are surrounded by low reticulate scarps marked by small angular reentrants. These scarps seem to be partly controlled by local structure and to be in an early stage of erosional retreat. Of importance with regard to the problem of scarp retreat on Io is the presence of the small white diffuse 'splotches' that emanate from an eroded section of one of the lower scarps. Earth-based spectra, along with IRIS data, suggest that they are frozen SO₂ on the ground^{4,9,10}. Similar features are present at the bases of most of the scarp types recognised on Io, and they are apparently related to alcoves and notches within the scarps themselves. Another significant relationship near the upper centre of *Plate 3* is the dark, 100-km long digitate flow that comes from the base of the lowest scarp in the stratigraphic sequence. This flow, unrelated to an observable caldera, is evidence of the presence of liquid material at shallow depth beneath the layered terrains.

Interpretations

Preliminary calculations indicate that ion bombardment from Jupiter's radiation field and consequent sputtering effects can remove only about 100 m of material in about 4 × 10⁸ yr, an insignificant rate in view of the resurfacing rate⁵. The scarp erosion patterns described here can be explained more realistically by internal sapping and the release of either volatiles or liquids to the surface. On Earth, where water is the dominant agent of erosion, spring sapping combined with mass wasting has an important role in the development of alcoves and reentrants in the cliffs of arid regions¹². Carr² has proposed an internal mechanism for Mars whereby water stored in the subsurface can be liquified and on bursting out on the surface will contribute to scarp retreat and even produce the fluvial features observed. No collected runoff or eolian activity is required in this model to account for scarp retreat.

Unlike Mars, we cannot appeal to ground ice and subsurface water to explain the fretted terrain. Io's surface is rich in sulphur and sulphur compounds and the large expanses of whitish terrain seen in the colour reconstructions probably consist either of frozen SO₂⁸⁻¹⁰ or a very white form of sulphur¹³ which has condensed out of volcanic exhalations.

Smith *et al.*⁷ have shown that SO₂ and sulphur should be solid on Io down to a depth of the order of 1 km or less. Below this level a phase change occurs with SO₂ becoming liquid and sulphur remaining solid. From a depth of about 2 km or less on down to a maximum depth of perhaps 4 km, sulphur is molten⁷. These relationships provide a basis for understanding the erosional scarps and the plumes found at their base as well as those seen in caldera walls and along fresh fault scarps. The subsurface isothermal trends marking the sulphur and SO₂ phase boundaries should approximately parallel the surface topography, except in the immediate vicinity of active volcanic centres. The edges of scarps would then represent the zones of easiest release for pressurised subsurface volatiles and liquids, as the layers described are of sufficient thickness to expect phase changes at depth behind the faces of the scarps. Thus, an artesian pressure would drive liquid SO₂ or sulphur to escape through cracks or fissures along the edges of boundary scarps.

A definite flow of molten sulphur from the base of a scarp might be expected to resemble the dark, 100-km long digitate flow referred to previously. The escape of SO₂ would be more dramatic. At artesian pressures of 15–30 bar, the liquid, close to its freezing point (198 K) is forced outwards along the escape

conduit until, near the surface, the pressure drops to the triple point (1.65×10^{-2} bar). We can follow the energetics using an isentropic path in the temperature–entropy diagram for SO_2 given by Smith *et al.*⁷. At the triple point, the liquid begins to crystallise and the system expands isothermally and isobarically with the formation of SO_2 vapour. An expanding effervescent mixture will now form, with each unit volume of liquid producing 5,000 times its volume of vapour. If the vapour and crystallising liquid remain in contact, the solid–fluid mixture will continue to expand until reaching the ambient atmospheric pressure ($\sim 10^{-7}$ bar) at or just outside the vent. On reaching the ambient pressure, the SO_2 solid–vapour mixture will have a total vapour content of $\sim 20\%$ by mass and an exit velocity of a little more than 350 ms^{-1} . The maximum range that can be reached by SO_2 snow issuing from a vent at this velocity is about 70 km for trajectories inclined $\sim 45^\circ$ to the horizontal. Depending on the orientation of individual vents, however, much of the distributed SO_2 snow would be expected to fall within a few tens of kilometres of the scarps. This is completely consistent with the scale of the diffuse white features seen in Plate 3 and we propose

that artesian SO_2 volatilisation, solidification and ballistic deposition is the source of these features.

The most likely initial escape route would be along a fresh boundary fault. This type of scarp, which is usually relatively straight and smooth surfaced, would rapidly become scalloped, notched and irregular in appearance, due to the energetic venting of SO_2 gas and snow along its face. This sapping process would work together with slumping and other mass wasting processes to create the erosional patterns observed. The release of the aquifer would continue until local equilibrium was attained. Recharge of SO_2 in the aquifer should cause further venting and continued erosion of the scarps. This process of venting and sapping would continue until the SO_2 is locally depleted (for example, the inselbergs) in which case further cliff retreat could occur only by mass wasting working alone. We thus propose that the fretted terrains on Io and those on Mars are formed by analogous sapping mechanisms in which liquid SO_2 provides the analogue to H_2O in terrestrial (and martian) aquifers.

Received 19 July; accepted 3 August 1979.

1. Sharp, R. P. *J. geophys. Res.* **78**, 4063–4072 (1973).
2. Carr, M. H. *J. geophys. Res.* (in the press).
3. Smith B. A. *et al. Science* **204**, 951–971 (1979).
4. Pearl, *et al. Nature* **280**, 755–758 (1979).
5. Johnson, T. V., Cook II, A. F., Sagan, C. & Soderblom, L. A. *Nature* **280**, 746–750 (1979).
6. Strom, R. G., Terrielle, R. J., Masursky, H. & Hansen, C. *Nature* **280**, 733–736 (1979).
7. Smith, B. A., Shoemaker, E. M., Keiffer, S. W. & Cook, A. F. II *Nature* **280**, 738–743 (1979).
8. Kumar, S. *Nature* **280**, 758–760 (1979).
9. Fanale, F. P. *et al. Nature* **280**, 761–763 (1979).
10. Smythe *et al. Nature* **280**, 766 (1979).
11. Carr, M. H. *Nature* **280**, 729–733 (1979).
12. Lucchita, B. K. *J. Res. U.S. Geol. Surv.* **16**, 128–137 (1977).
13. Sagan, C. *Nature* **280**, 750–753 (1979).

The role of SO_2 in volcanism on Io

B. A. Smith*, E. M. Shoemaker†, S. W. Kieffer† & A. F. Cook II‡

* Department of Planetary Sciences, University of Arizona, Tucson, Arizona 85721

† US Geological Survey, Flagstaff, Arizona 86001

‡ Smithsonian Astrophysical Observatory, Cambridge, Massachusetts 02138

Io and Earth are the only planetary bodies known to be volcanically active; the energetics of the eruptive plumes on Io have important structural implications and are closely linked with the presence of sulphur and SO_2 .

VOLCANIC plumes observed on Io by Voyager 1 reach heights between 70 and 280 km^{1,2}, implying vent ejection velocities ranging from several hundreds of metres to nearly a thousand metres per second (Table 1). At least eight vents were active at the time of Voyager 1 encounter and both recent and current activity of others seems likely². Although evidence for volcanism is found on the Moon, Mars and Mercury, Io is the only planetary body other than the Earth known to have active volcanism, and the impressive rate at which volcanic material is being deposited on the surface of this highly active planet has profound implications for its bulk chemical evolution³. Here we examine the energetics of the eruptive plumes and the implications for the structure, the ground fluid 'hydrology' and hydrological cycle of Io's crust.

The driving energy for Io's intense volcanism seems likely to be the model proposed by Peale *et al.*⁴ where tidal energy is dissipated in a thin crust surrounding a completely molten interior. In this model, runaway melting takes place within a small molten core and is stabilised only when the solid mantle or crust becomes sufficiently thin to conduct away all of the tidally generated heat. Because this mechanism is sustained by the forced eccentricity of Io's orbit which arises chiefly from perturbation by Europa, it probably existed throughout much of geologic time and should continue almost indefinitely. If we assume that the current resurfacing rates³ are applicable over geological time scales, Io's crust must have been recycled many times over while in contact, at its base, with a convective interior. Thus, early loss and eventual depletion of volatiles from the bulk composition of Io would be expected. In explaining current

volcanism on Io, we must consider those materials with vapour pressures sufficiently low to have survived aeons of crustal and internal recycling.

Fundamental to our analysis is the probable presence of SO_2 frost or adsorbed SO_2 on the surface of Io^{5,6} and the recognition that SO_2 is a prominent constituent of its atmosphere^{3,7–9}. Noting also the discovery of ionised S and O in the Io torus^{10,11}, the spectrophotometric suggestion of elemental sulphur on its surface^{12–15}, and the absence of both H_2O bands and the broad absorption features associated with iron-bearing silicates, we conclude that SO_2 is the volatile which drives pyroclastic volcanism on Io and that sulphur and SO_2 are the primary constituents of the uppermost crust and subjacent molten zone. Other constituents are certainly present (Na and K are observed to be escaping from Io), but probably occur in minor quantities.

The mechanism that drives violent gaseous eruptions in terrestrial volcanoes is the rapid expansion of volatiles, primarily H_2O and CO_2 , which may either be dissolved in the silicate magma or raised to high temperature ($\sim 1,500 \text{ K}$) by contact with magma intrusions. Although vent velocities of hundreds of metres per second have been observed in terrestrial volcanic eruptions¹⁶, interaction with the terrestrial atmosphere limits the heights of volcanic plumes to a few tens of kilometres at the most. This is not the case on Io, where venting gases abruptly enter a near-vacuum environment.

Limiting velocities of a fluid or fluidised system expelled from a volcanic vent can be estimated using a simple model of adiabatic decompression of the system as it moves upwards through a passageway connecting a chamber or arbitrary depth with the surface. Assuming reversible adiabatic flow and chemical equilibrium, mass flow up the volcanic conduit can be considered to be isentropic. For such a simplified process, the thermodynamic path of the erupting material during ascent is easily envisioned on a temperature–entropy phase diagram, shown in Fig. 1 for SO_2 .

Table 1 Heights of volcanic plumes on Io

Io volcanic plumes (no.)	Maximum height above the surface* (km)	Vent velocity (ms ⁻¹)
1	280	930
2	100	580
2	210	810
(UV component)		
3	70	490
4	95	570
5	80	520
6	80	520
7	120	630
8	70	490

* After Strom *et al.*².

Several qualitatively different eruption styles are possible, depending on the entropy of the system at the time the eruption begins. High velocities of the system are attained after a vapour phase is formed. This can occur when a fluid system rising from depth is decompressed below the critical point, for cases of high entropy, or when the boundary of the liquid+vapour field is reached, for cases of low entropy. It can also occur if the system remains in place until its temperature is raised (by intrusion of a hot mass) to the point where boiling begins.

If a system such as SO₂ has an entropy less than the critical point entropy when vapourisation begins, and if the vapour and liquid phases do not separate, then, on isentropic expansion during ascent of the system to the surface, vapour will continuously evolve as the system varies along an isentrope passing through the liquid+vapour field (see line ABC in Fig. 1) until the triple point (0.0165 bar for SO₂) is intercepted. At this point, remaining liquid in the system crystallises to a lower entropy solid, and vapour is added simultaneously, causing the system to expand isobarically and isothermally until all the liquid phase disappears. Further decompression along the same isentrope results in condensation of SO₂ ice from the vapour phase; complete expansion to 0 K would result in condensation of all the vapour.

Assuming complete thermodynamic equilibrium and steady-state flow, the specific kinetic energy available for fluid motion at any pressure or temperature on ascent of the system to the surface is the difference between the initial specific enthalpy, H_0 , and the enthalpy of the products, H minus the specific potential energy gained by rise in a gravity field. For one dimensional flow,

$$\frac{u^2}{2} = H_0 - H - \int g \, dh \quad (1)$$

where u is the velocity of the system; g the gravitational acceleration; and h the height of the system.

The enthalpy of the products is

$$H = xH_v + (1-x)H_{lors} \quad (2)$$

where H_v and H_{lors} are the enthalpies of the vapour, liquid or solid phases, and x is the mass fraction of vapour present; x is obtainable from the isobaric chord ratios in Fig. 1 or, numerically, from

$$x = \frac{S - S_{lors}}{S_v - S_{lors}} \quad (3)$$

where S , S_v , S_{lors} are the total entropy, and vapour, liquid, or solid entropies respectively.

As a specific example of flow at an entropy less than the critical point entropy, consider flow from a liquid SO₂ reservoir at 40 bar and 393 K (representing flow initiating from contact with pure sulphur on its liquidus temperature of 393 K at approximately 1.5 km depth). The entropy of the SO₂ system is 2.90 J g⁻¹ K⁻¹ (point A in Fig. 1) and its initial specific enthalpy is 530 J g⁻¹. Formation of vapour in the SO₂ system initiates upward movement of the two-phase (liquid+vapour) system and the material flows up the volcanic vent with continuous evolution of vapour and increasing velocity until the triple point is reached. At this point, the SO₂ constitutes a liquid-vapour fluidised system consisting of 45% (mass fraction) vapour and, neglecting small values of $\int g \, dh$, is moving at 430 m s⁻¹ (Table 2). If it is assumed that the vent discharges at pressures in

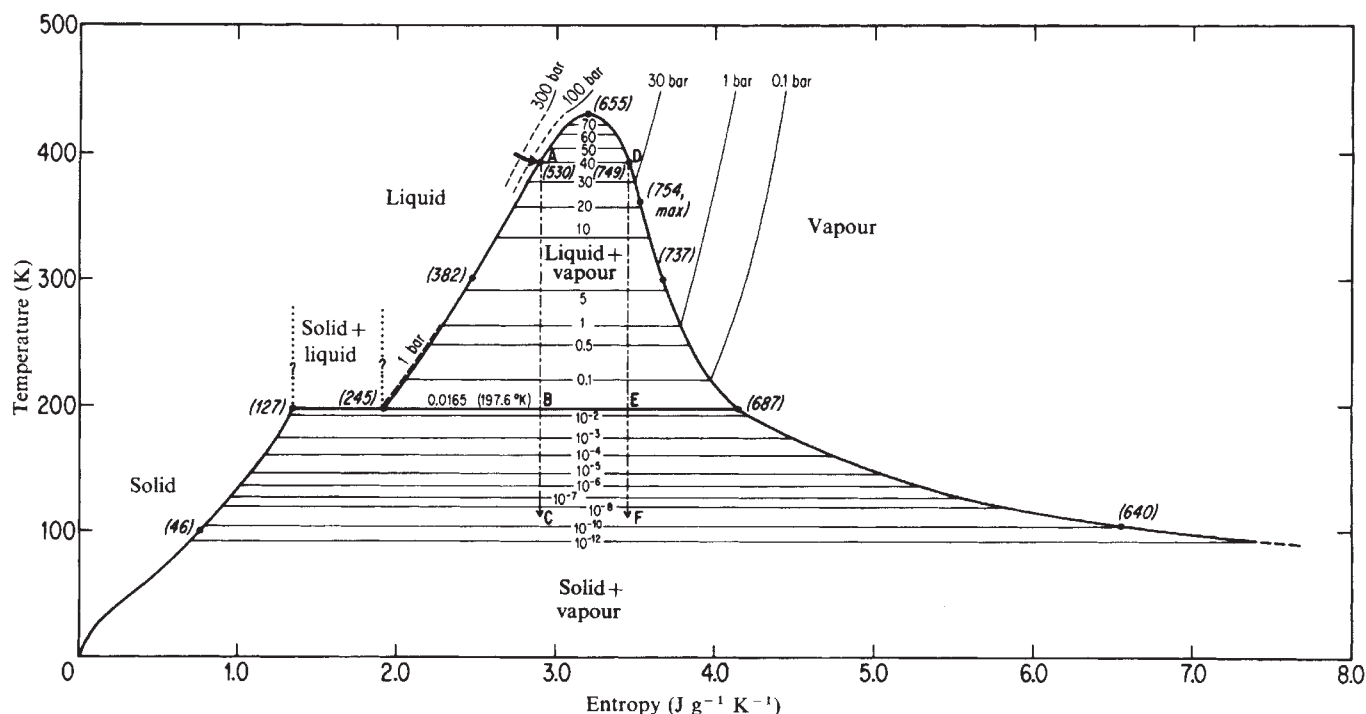


Fig. 1 Entropy-density relations in the SO₂ system. Heavy solid lines indicate the phase boundaries; heavy dashed or dotted lines are extrapolated beyond data sources; light lines represent isobars. Liquid-vapour equilibrium data from Braker and Mossman²⁶ adjusted to entropy equal 0.0 for the solid phase at 0 K by using the triple-point entropy and the enthalpy of fusion of Giauque and Stephenson²⁷. The solid sublimation curve was calculated from the heat capacities of Giauque and Stephenson by numerical integration. The vapour curve on the righthand side was calculated from heats of sublimation provided by D. D. Wagman (personal communication). The 1 bar and 30 bar isobars and all isobars in the liquid+vapour region are from vapour pressures listed by Braker and Mossman; the 0.1, 100, and 300 bar isobars are from data of Canjai and Manning²⁸. Isobars in the solid+vapour region are interpolated from vapour-pressure data provided by Wagman. Heavy arrow is pressure-temperature path of the sulphur liquidus, which intersects the SO₂ liquid-liquid+vapour phase boundary at point A. Isentropes ABC and DEF, shown with dot-dash lines, are discussed in text. Specific enthalpies are shown in parentheses.

Table 2 Mass fractions of vapour and velocities of SO₂ fluidised systems expanding isentropically from an initial state at 393 K and 40 bar

Pressure (bars)	Low entropy case (expansion from A in Fig. 1)		High entropy case (expansion from D in Fig. 1)	
	x	u (m s ⁻¹)	x	u (m s ⁻¹)
30.1	0.11	—	0.95	124
14.4	0.25	93	0.89	246
8.0	0.32	149	0.86	320
1.17	0.40	260	0.79	461
0.107	0.44	370	0.73	571
0.020	0.44	421	0.70	625
0.0165	0.55*	433*	0.75*	637*
1.9 × 10 ⁻³	0.52	488	0.70	689
9 × 10 ⁻⁵	0.48	550	0.64	752
7 × 10 ⁻⁷	0.44	610	0.58	815
8 × 10 ⁻⁹	0.41	680	0.52	870
4 × 10 ⁻¹²	0.35	737	0.40	965
0.00	0.00	1030	0.00	1224

*Calculated at the triple point when the liquid phase has just disappeared.

the range 10⁻⁷–10⁻¹² bar at the ground surface (that is, that the isobars of the atmosphere are not displaced large distances upward over the discharging vent) then, from the temperature–entropy diagram (Fig. 1), it can be concluded that the triple point is reached at depth before the fluidised system emerges from the surface. On the other hand, if the surface isobars are greatly displaced by the flow (the eruption plume ‘punches through’ the Io atmosphere), the triple point might be reached above the ground surface. At the triple point, the liquid droplets crystallise as the system expands isothermally and isobarically with an increase in vapour content to 55%. Further expansion is accompanied by condensation of SO₂ ‘snow’ from the vapour phase; at the time of exit to surface pressure, the vapour content of the system has decreased to 45–40% (depending on the surface pressure assumed) and, for the case considered, the velocity at the vent is 600 to 700 m s⁻¹. As the plume expands above the surface, SO₂ snow continues to form and the column continues to accelerate; if the plume were to break through the Io atmosphere and expand to approximately zero pressure and 0 K, the enthalpy converted to kinetic and potential energy would be equivalent to a velocity of 1.03 km s⁻¹, if the energy released were represented entirely by the kinetic energy term.

A different process occurs if the SO₂ system has an entropy higher than the critical point entropy on initiation of the eruption. Such a higher entropy could be attained either by (1) heating the SO₂ by incorporation of hot (sulphur) fragments or lapilli; or (2) segregation in the reservoir of the higher entropy vapour from the liquid SO₂, for example, across the 40 bar isobar in the temperature–entropy diagram; or by (3) intrusion of the heat source to shallow depth where the pressures are <40 bar. For any of these cases, the two-phase field is entered at the boundary with the vapour field rather than at the boundary with the liquid field, and isentropic expansion follows a path such as DEF shown in Fig. 1. The initial ascent and expansion of the system are accompanied by condensation of SO₂ liquid from the vapour; at the triple-point the condensed liquid freezes and further ascent is accompanied by condensation of solid from the vapour.

As a specific example of flow at an entropy higher than the critical point entropy, consider the same reservoir as discussed previously at 393 K, 40 bar, but with the vapour phase continuously segregated from the lower-entropy liquid (corresponding to the vapour side of the 40 bar isobar in Fig. 1). The entropy of the vapour is 3.46 J g⁻¹ K⁻¹ (point D) and its initial specific enthalpy is 749 J g⁻¹. As the vapour accelerates upwards, liquid is continuously condensed until triple point conditions are reached (point E). At this point, the SO₂ system consists of 70% vapour and is moving at 625 m s⁻¹ (again neglecting changes in potential energy). At the triple point the liquid freezes, the vapour content increases to 75%, and the system expands isobarically. Further ascent of the system is accompanied by condensation of SO₂ ‘snow’ from the vapour phase. At the time

of exit to surface pressure, the vapour content of the system has decreased to 55–50% and the exit velocity is in the range 850–950 m s⁻¹ (for surface pressures of 10⁻⁷–10⁻¹² bar). The theoretical maximum velocity obtainable (by expansion to zero pressure) is 1,224 m s⁻¹.

If acceleration of the fluidised system during expansion above the surface of Io is neglected, the maximum height above the surface reached by a parcel of the system leaving a vent at some exit velocity, *u*, will depend on the direction of its trajectory relative to the local vertical, which in turn is determined by the geometry of the vent itself. If we assume that the trajectory is purely ballistic, that most of the gas and entrained solids are ejected directly toward the local vertical, and that we can neglect the gravitational perturbations due to Jupiter, a given exit velocity, *u*, can be related to a maximum height, *h*, above the surface reached by the ejected material by

$$h = 2GM (V^2 - u^2)^{-1} - r_0 \quad (4)$$

where *V* is the two-body velocity of escape, *r*₀ the radius of Io, *M* the mass of Io, and *G* the gravitational constant. Taking 8.91 × 10²² g and 1,820 km for the mass and radius of Io and ignoring perturbations due to Jupiter, the velocity of escape from the surface of Io is 2.55 km s⁻¹. Assuming the plume expands to pressures of 10⁻⁷–10⁻¹² bar, the maximum heights reached by the decompression of pure SO₂ from the starting conditions at 393 K and 40 bar are 130–170 km for the low entropy case and 230–310 km for the high entropy case.

When computing actual escape trajectories, we cannot ignore the effects of Jupiter. For example, particles ejected vertically from the surface of Io directly towards or away from Jupiter with a velocity of 2.28 km s⁻¹ would reach the radius of the co-linear lagrangian points (*L*₁ and *L*₂) ~8,700 km above the surface and thus escape into circumjovian space. A more comprehensive discussion of this three-body problem as applied to Io is given by Smyth and McElroy¹⁷. When treating the near space within 300 km of Io's surface, however, the magnitude of Jupiter's perturbations remains <0.003 and the effects on equation (4) are negligible.

A slightly more stringent requirement on the vent exit velocities than the height of plumes is given by the maximum distance to which material has been deposited on the surface of Io by the plumes. Each deposit records the maximum range over the life of the plume. In the case of Plume 1, the most distant limit of the recognisable deposit is 700 km (22°) from the eruptive centre. The minimum velocity at which material can be ejected to this range on a simple ballistic trajectory is given by

$$u = \left(\frac{2r_0g_0}{\cot(\phi/2) + 1} \right)^{1/2} \quad (5)$$

where *g*₀ is the gravitational acceleration at the surface (179 cm s⁻²); and *φ* the range in planetocentric angular distance from the vent. For a range *φ* of 22° for the limit of the deposit of Plume 1, *u* is 1.03 km s⁻¹.

Significant departures from purely ballistic trajectories of material in the plumes are caused by expansion of the gas in the plumes and probably also by shocks. Expansion in the plumes provides a moderate increment of velocity after exit from the vent, whereas shocks influence the flow in the gas and details of the shapes of the plumes and trajectories of initially entrained particles¹⁸. We can conclude, however, that SO₂ gas and condensed particles can be accelerated to velocities required to explain the highest of the observed plumes, starting from temperatures close to the sulphur liquidus, so long as the plumes are nearly free of entrained initially cold material. A temperature slightly higher than the melting point of monoclinic sulphur at 40 bar (393 K) may be required to explain the most distant deposits of Plume 1.

Addition of cold solid material to the fluidised system expanding up the vent, by spalling of the walls of the vent, for instance, will decrease the velocity of the system. If all this material is entrained, the kinetic energy of the fluidised system

must be distributed through a larger mass. If the mass of the system is doubled, for example, the velocity will decrease by a factor of $\sqrt{2}$. On the other hand, if liquid sulphur is entrained, the decrease in velocity is very much less than for cold wall rock. This is because the enthalpy of the sulphur droplets becomes converted to kinetic energy as the fluidised system expands and cools. Entrained sulphur is probably reduced to particles comparable in size to volcanic ash produced in violent gaseous eruptions in terrestrial volcanoes of the order of 1 mm or less. Drops of this size will very quickly come to approximate thermal equilibrium with the gas, releasing both the heat of fusion, $\sim 38 \text{ J g}^{-1}$, as the drops freeze, and about $0.75 \text{ J g}^{-1} \text{ K}^{-1}$, as the frozen sulphur ash cools to temperatures of the order of 100 K. The total specific enthalpy released by cooling liquid sulphur to 100 K is about 260 J, which is about half of the specific enthalpy released by adiabatic expansion of the entraining SO_2 gas. Hence, fairly large mass fractions of initially liquid sulphur can be entrained in the eruptive column without seriously decreasing the vent exit velocity of the plume.

Observations made by Voyager I are consistent with the interpretation that elemental sulphur is a common constituent of the surface of Io. Abundant flows of low relief, ranging in colour from red to yellow, radiate from many caldera-like features^{19,20}. These flows have lengths of up to several hundred kilometres, and many apparently have very low gradients. They were formed, therefore, by fluids of viscosity comparable to or less than the viscosity of lunar basaltic lavas, $\sim 10^2\text{--}10^4 \text{ P}^{21}$. Molten sulphur at all temperatures is a fluid of appropriately low viscosity²⁰. Moreover, because of the low melting point of sulphur, the rate of heat loss from sulphur flows would be two orders of magnitude lower than for silicate lava flows, which further facilitates the development of comparatively great flow length.

Many calderas are occupied by material of very low albedo resembling lava lakes^{1,19,22}. In one case, a black horseshoe-shaped 'lava lake' $\sim 200 \text{ km}$ across (Plate 1, p. 782) has a model-derived temperature about 150 K above that of the surrounding surface²³. Thermal observations suggest that this feature is an active lake of molten sulphur with a partial, very thin, chilled crust. Of special interest is that although this 'lake' lies within the fallout area of Plume 2, no fallout deposit is recognisable on its surface. The albedo of the 'lake' is appropriate for molten sulphur at temperatures in the range of 500 K and above²⁰. The black colour of the liquid may be partly preserved in chilled crusts by quenching²⁰. We interpret all the black features resembling lava lakes distributed across the surface of Io as active or recently active lakes of molten sulphur.

Except at the 'lava lakes' and possibly other active lava surfaces, the surface temperature of Io is well below the freezing point of SO_2 (198 K at 1 bar). The widely distributed white patches, illustrated elsewhere¹, could well be optically thick deposits of SO_2 frost, although a very white form of elemental sulphur has also been suggested²⁰. SO_2 may also be adsorbed on sulphur (or other constituents) on the yellow parts of Io's surface. Any plume-supplied SO_2 , leading to a surface pressure somewhat in excess of the mean daytime (130 K (ref. 23)) vapour pressure of solid $\text{SO}_2 \sim 1.4 \times 10^{-7} \text{ bar}$ (Fig. 3), would result in precipitation of SO_2 frost and the build up of frost deposits that are stable through the day. Frost would tend to accumulate preferentially in areas that remain coldest during the diurnal temperature cycle. At a given latitude the areas that remain coldest are those with the highest albedo—the white patches themselves. This positive feedback mechanism may explain the patchy distribution of the frost and the fact that lava flows recently extruded onto the surface apparently remain frost-free (at least during the day).

Models of the crustal structure of Io, which are necessary for a theoretical understanding of the dynamics of the plumes, can be derived from an estimate of mean density of the satellite, estimates of the rate of heat production by tidal dissipation, and the interferences about the surface composition. The mean density of Io, 3.53 g cm^{-3} (ref. 1), suggests that the satellite is

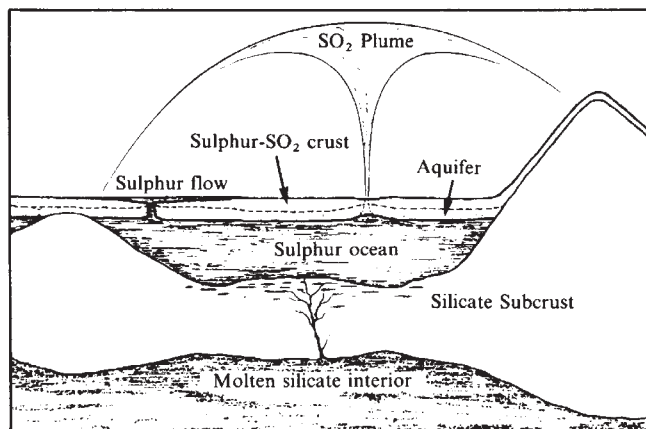


Fig. 2 Proposed crustal structure for Io. Heights and thicknesses are not drawn to scale. The uppermost layer, which also covers all solid silicate topography, is composed of a mixture of solid SO_2 and sulphur. The subjacent layer is a porous solid sulphur aquifer through which liquid SO_2 flows freely. Below this is a molten sulphur 'ocean' which rests on a solid silicate subcrust overlying a uniformly melted silicate interior. Where the upper crust is less buoyant, molten sulphur will flow over the surface. A plume is created when liquid SO_2 pours into an open vent left by a receding molten sulphur intrusion. Intruded silicate sills build up at the bottom of the sulphur ocean.

composed predominantly of ferromagnesian silicates and that it has a substantial iron or iron sulphide core. The model for Io proposed by Peale *et al.*⁴ before the Voyager discovery of volcanism contained a uniformly melted interior surrounded by a 20 km thick silicate crust with an implied surface heat flux of $\sim 500 \text{ erg cm}^{-2} \text{ s}^{-1}$. More detailed models of the crust depend critically on the assumed ratio of free sulphur to silicates in the crust. As discussed by Carr¹⁹, sulphur dissolved in silicate magmas exsolves as the magmas rise to the surface and, in terrestrial volcanoes, condenses on the surface or in near surface voids. This condensed sulphur is then remobilised as a melt if it is buried and reheated, a process which has occurred on Earth where sulphur lava flows have been noted. Buildup of the Io crust by repeated extrusion of sulphur-saturated silicate lavas would result in continued enrichment of sulphur in the upper part of the crust by remobilisation and upward migration of the free sulphur both as a gas and as a melt.

Carr¹⁹ suggests that both sulphur and silicate lava flows are present around many caldera-like eruptive centres on Io. If this is so, the solid crust of Io consists predominately of silicates, with a subordinate amount of sulphur concentrated in the upper part. However, injection of a column of silicate magma into a solid sulphur upper crust would quickly lead to melting of a volume of sulphur larger than the volume of the silicate column; the higher density silicate magma could not then rise buoyantly through the fluid sulphur unless it were expanded by vesiculation to a density less than that of molten sulphur ($\sim 1.8 \text{ g cm}^{-3}$).

We prefer a model in which the upper crust consists largely of elemental sulphur and SO_2 and overlies a subjacent layer of molten sulphur possibly several kilometres thick (Fig. 2). In the upper crust, sulphur is present both as lava flows and as interstratified pyroclastic deposits. The rate at which temperature increases with depth in the upper crust depends both on the local coefficient of thermal conductivity of the sulphur and the frozen SO_2 and the value of the net outward heat flux. Of the two, the latter is by far the more difficult to estimate. In our model, we use a conservative value for the heat flux generated by tidally dissipated energy in the silicate lower crust, $100 \text{ erg cm}^{-2} \text{ s}^{-1}$. Of this total, using resurfacing rates estimated by Johnson *et al.*³, $25 \text{ erg cm}^{-2} \text{ s}^{-1}$ escapes through volcanism, leaving $75 \text{ erg cm}^{-2} \text{ s}^{-1}$ to be conducted upward through the crust. The lower crust (a few km to $\sim 20 \text{ km}$ depth) is composed mainly of a complex of silicate sills that have been intruded successively at the base of the molten sulphur layer. These intrusives are more

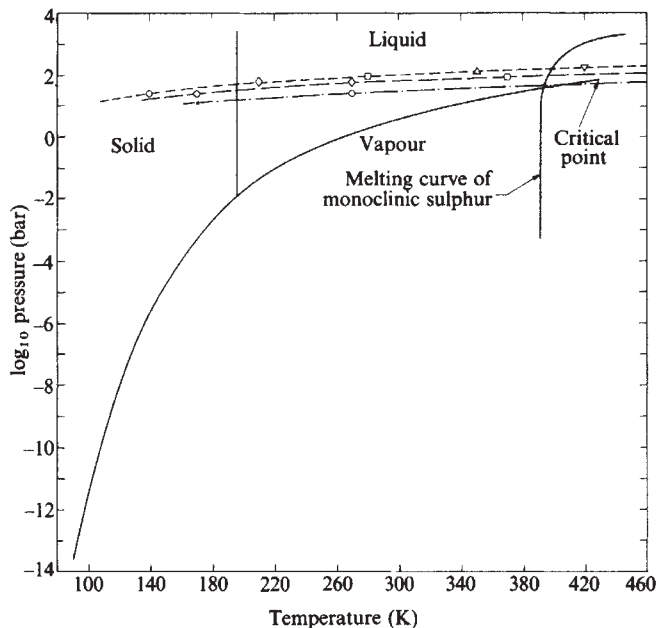


Fig. 3 Pressure-temperature phase diagram for SO_2 showing the relationship of three assumed geotherms on Io to the boundaries of the phase fields. Boundaries of phase fields are derived from data given by Braker and Mossman²⁶, Giaque and Stephenson²⁷ and Wagman. The geotherm 200 K km^{-1} is believed to be close to the actual pressure-temperature profile in the upper (sulphur) crust of Io. Depths plotted on the geotherms are calculated for densities of 1.45 g cm^{-3} up to 1 km and 1.89 g cm^{-3} below 1 km (corresponding to a porous sulphur- SO_2 permafrost zone to 1 km and a zone of porous sulphur saturated with liquid SO_2 below 1 km). $\odot$, 0.5 km; $\square$, 1 km; $\diamond$, 2 km; $\square$, 3 km; $\triangle$, 4 km; ∇ , 5 km; ----, 70 K km^{-1} ; ---, 100 K km^{-1} ; —, 200 K km^{-1} .

and more thermally metamorphosed towards the base of the lower crust where, again, complete remelting defines the base.

Recycling of the sulphur upper layer is independent of the cycling of the silicate lower layer. Temperatures of sulphur melts erupting at the surface will tend to be close to the melting point of sulphur.

The limiting thickness to which the sulphur upper layer can evolve is that at which the pore pressure at the base of the sulphur layer is just equal to the mean sulphur vapour pressure of the silicate magmas intruded at its base. Studies by Haughton *et al.*²⁴ suggest that silicate magmas with sulphur vapour pressure high enough for the sulphur upper layer to evolve to several kilometres thickness (S_2 vapour pressure ~ 100 bar) must have very low contents of FeO.

SO_2 frost precipitated on the surface of Io tends to be buried under pyroclastic fallout from the eruptive plume systems, and, if sufficiently insulated by pyroclastic deposits, it can also be buried beneath sulphur flows. Thus, as volcanic material accumulates on the surface, some of the frost deposits are trapped and carried down in the volcanic cycling of the crust. As temperatures rise with increasing depth of burial, the trapped SO_2 probably migrates to some extent and recrystallises in irregularly distributed bodies of ground ice.

Assuming the nominal heat flux given above and a temperature-dependent, thermal conductivity of $3\text{--}5 \times 10^{-3} \text{ W cm}^{-1} \text{ K}^{-1}$ for the permafrost zone, the 198 K isotherm (freezing level of SO_2) will lie at depths between 500 and 1,000 m; depths to this isotherm will be greatest at the poles and least at the equator. Below this isotherm the trapped SO_2 is liquid and forms an interstitial ground fluid that, if sufficiently abundant, tends to saturate all permeable strata. The SO_2 ground fluid freezes out wherever it reaches the 198 K isotherm and tends to seal the base of the overlying SO_2 permafrost zone to produce an effective 'aquaclude'.

The base of the SO_2 liquid zone will be defined either by contact with molten sulphur or by intercept of the local geotherm of Io with the SO_2 vapour field (Fig. 3). For thermal

gradients of the order of 200 K km^{-1} or somewhat steeper consistent with the heat flow and thermal conductivity we have adopted, the SO_2 liquid zone will be $\sim 500\text{--}1,000$ m thick, and it will be underlain by a thin SO_2 vapour zone. If the total abundance of SO_2 in the crust is high enough, the SO_2 liquid zone may be saturated or nearly saturated throughout.

If the SO_2 liquid zone is saturated, it will be an artesian system, driven by the pressure of liquid sulphur or SO_2 vapour at the base of the zone. In this case, liquid SO_2 will flow towards any locus where fluid escapes and where there is a corresponding pressure drop⁹. If many such loci of fluid escape are distributed over the surface of Io, the regional flow of SO_2 ground fluid will tend to be up the pressure gradient, towards the shallowest parts of the 198 K isotherm. Hence, the regional flow will tend to be from the polar regions to the equator, possibly from the regions of white patches to frost-free regions and, most importantly, from regions of average thermal gradient towards active plume vents.

All but one of the known eruptive plumes on Io occur at latitudes $\leq 30^\circ$ (ref. 2). Assuming that the plumes are driven by SO_2 , this distribution suggests that the SO_2 liquid zone is, in fact, nearly saturated and that regional artesian flow occurs from the polar regions to the equatorial regions. If only a small fraction of the SO_2 liquid zone were saturated, most or all of this zone at low latitudes might be above the regional SO_2 liquid table and insufficient SO_2 would be available to form the eruptive plumes.

The melting point of sulphur is probably reached at depths of the order of 1,500 m below the surface. In places where convective flow of the SO_2 ground fluid can occur, the thermal gradients will be less steep and the top of the molten sulphur layer may be deeper. Where the sulphur is thick enough, the molten sulphur constitutes a sulphur 'ocean' or 'thiosphere'²⁰ extending downwards to the lower silicate crust, possibly several kilometres below. The solubility of SO_2 in liquid sulphur is ~ 0.02 (mass fraction): thus SO_2 may also be transported over global distances by currents within the sulphur ocean.

The upper crust of solid sulphur and solid and fluid SO_2 rests buoyantly on the sulphur ocean. The upper crust is buoyant because of pore space in the SO_2 permafrost zone, probably on the order of several tens of per cent, and the admixture of low density solid SO_2 ($\sim 1.6 \text{ g cm}^{-3}$), in the solid sulphur ($\sim 2.0 \text{ g cm}^{-3}$), and because of the low density of the liquid SO_2 ($\sim 1.5 \text{ g cm}^{-3}$) that fills the pore space in the SO_2 liquid zone. As the SO_2 liquid zone is probably sealed at the base of the permafrost, pore pressure in the liquid SO_2 can approach the lithostatic pressure. Hence, if the SO_2 liquid zone is saturated, collapse of pores is retarded until the SO_2 vapour zone is reached. Wherever the base of the sulphur upper crust exceeds the density of liquid sulphur, (1.8 g cm^{-3}) as result of loss of pore space, it will tend to founder into the sulphur ocean.

The overall buoyant freeboard of the sulphur crust is maintained by delamination or foundering of the base but probably remains small. Wherever the freeboard is negative, liquid sulphur can rise to the surface to form a sulphur volcano. Rapid rise of sulphur along a new fissure will result in chilling of the sulphur against the fissure walls. A layer of frozen sulphur along the chilled margins probably seals the fissure against inflow of liquid SO_2 . In most cases, cessation of sulphur magma flow at an immature vent probably results in complete freezing of sulphur in the vent and no loss of SO_2 fluid at the surface.

At more mature conduits, where prolonged flow of sulphur has heated the vent walls (perhaps those vents that feed lava lakes) there may be a different denouement of a sulphur volcano. When local foundering at the base of the sulphur crust occurs, the crust will bob up buoyantly, and the level at the top of the liquid sulphur column will abruptly subside relative to the surface. Because the top of the liquid SO_2 zone surrounding a mature vent may be near the surface, withdrawal of the sulphur column allows the SO_2 to flow into the vent, initiating a 'phreatic' eruption. By this means a plume is born. The suggested mechanism for initiating the plumes is analogous to the mechanism of rare phreatic eruptions at Kilauea, Hawaii, which

occur after major withdrawals of lava in the caldera that allow groundwater to flow into the vent²⁵. Where the flow of SO₂ ground fluid is high enough, the top of the liquid sulphur column will be eroded by the expanding fluidised SO₂ system, and the base of the expanding system may descend deep into the sulphur crust. The plume may be extinguished when the liquid SO₂

available to drain into the vent is exhausted or, more likely, when the rate of upwelling of liquid sulphur in the vent exceeds the loss to the fluidised column at the top. This could occur after the SO₂ liquid table around the vent has been deeply drawn down by prolonged discharge. The rising liquid sulphur column will then shut off the SO₂ flow.

Received 19 July; accepted 9 August 1979.

1. Smith, B. A. *et al. Science* **204**, 951–972 (1979).
2. Strom, R. G., Terrile, R. J., Masursky, H. & Hansen, C. *Nature* **280**, 733–736 (1979).
3. Johnson, T. V. *et al. Nature* **280**, 746–750 (1979).
4. Peale, S. J. *et al. Science* **203**, 892–894 (1979).
5. Fanale, F. P., Hamilton Brown, R. & Cruikshank, D. P. *Nature* **280**, 761–763 (1979).
6. Smythe, W. D. *et al. Nature* **280**, 766 (1979).
7. Kumar, S. *Nature* **280**, 758–760 (1979).
8. Pearl, J. *et al. Nature* **280**, 755–758 (1979).
9. McCauley, J., Smith, B. A. & Soderblom, L. A. *Nature* **280**, 736–738 (1979).
10. Broadfoot, A. L. *et al. Science* **204**, 972 (1979).
11. Krimigis, S. M. *et al. Science* **204**, 998–1002 (1979).
12. Wamsteker, W. *et al. Icarus* **23**, 417 (1974).
13. Nash, D. B. & Fanale, F. P. *Icarus* **81**, 40 (1977).
14. Nelson, R. & Hapke, B. *Icarus* **36**, 304 (1978).

15. Fanale, F. P. *et al. Science* **186**, 922 (1974).
16. Self, S. *et al. Nature* **277**, 440 (1979).
17. Smyth, W. H. & McElroy, M. B. *Planet Space Sci.* **25**, 415 (1977).
18. Cook, II, A. F. *et al. Nature* **280**, 743–746 (1979).
19. Carr, M. H. *et al. Nature* **280**, 729–733 (1979).
20. Sagan, C. *Nature* **280**, 750–753 (1979).
21. Cukierman, N., Tutts, P. N., & Uhmman, D. R. *Geochim. cosmochim. Acta Supl* **3**, 2619 (1972).
22. Masursky, H. *et al. Nature* **280**, 725–729 (1979).
23. Hanel, R. *et al. Science* **204**, 972–976 (1979).
24. Haughton, D. R. *et al. Econ. Geol.* **69**, 451 (1974).
25. Finch, R. H. *Pacif. Sci.* **1**, 237 (1947).
26. Braker, W. & Mossman, A. L. *Matheson Gas Data Book*, 5th edn, 513 (C. Rutherford, New Jersey, 1971).
27. Giauque, W. F. & Stephenson, C. C., *J. Am. chem. Soc.* **60**, 1389 (1938).
28. Canjai, L. N. & Manning, F. S. *Thermodynamic Properties and Reduced Correlations for Gases*, 169 (Gulf Publishing Co., Houston, 1969).

Dynamics of volcanic plumes on Io

A. F. Cook*, E. M. Shoemaker† & B. A. Smith‡

* Center for Astrophysics, Harvard College Observatory and Smithsonian Astrophysical Observatory, Cambridge, Massachusetts 02138

† US Geological Survey, Flagstaff, Arizona 86001

‡ Department of Planetary Sciences, University of Arizona, Tucson, Arizona 85721

Ballistic and aerodynamic models are proposed to explain the volcanic plumes on Io, with particular reference to Plumes 1 and 3 which seem to have the same origin.

VARIOUS clues suggest that aerodynamics play a role in the volcanic plumes on Io which were revealed by the cameras on Voyager 1¹. These are: (1) the formation of a more or less circular ring or double ring about the vent, the region of the ring being a bland one obviously covered over by frost or crystals or entrained particles from the plume; (2) the filamentary structure of the upper regions of Plume 1 (Plate 6b, p. 788) as shown in a picture taken near the limb; and (3) abrupt bends in each of the individual dark filaments along trajectories in Fig. 1 which shows Plume 3 taken as it was becoming visibly active on the disk.

The ballistic model

We assume that the rate of flow of gas is sufficiently small from our volcanic vent so that gas and solid particles become completely decoupled over a distance small compared with the maximum height reached by the particles and remain decoupled. In that case a convenient model starts all particles through a circular orifice with a common velocity diverging from a common point below the centre so that the velocity vectors uniformly fill a cone with a vertical axis and opening out toward the zenith. We shall show that this model does not fit the observed structure of the plumes.

Let r denote radial distance (Fig. 2), z the height above the surface, t the time, z_0 the depth of the vertex of the launch cone, v_0 the effective launch velocity at that depth, and t_0 the effective epoch of passage through the vertex. We neglect the curvature of Io's surface and the variation of its acceleration of gravity g , with height. The equations of motion and initial conditions are:

$$\left. \begin{aligned} d^2z/dt^2 &= -g, & d^2r/dt^2 &= 0, & dz/dt|_{t=t_0} &= v_0 \cos i, \\ dr/dt|_{t=t_0} &= v_0 \sin i, & z|_{t=t_0} &= -z_0 & \text{and} & r|_{t=t_0} = 0 \end{aligned} \right\} \quad (1)$$

where i denotes the launch angle measured from the zenith.

Elimination of the time and solution for the two values of i which will dispatch particles through a given position (r, z) yields:

$$\cot i = \frac{v_0^2}{gr} \left\{ 1 \pm \left[1 - \left(\frac{gr}{v_0^2} \right)^2 - 2 \frac{g(z_0 + z)}{v_0^2} \right]^{1/2} \right\} \quad (2)$$

The two solutions coalesce at an upper envelope at which the square root vanishes

$$z_M = -z_0 + \frac{v_0^2}{2g} - \frac{gr^2}{2v_0^2}, \quad \cot i_M = \frac{v_0^2}{gr} \quad (3)$$

where the subscript M denotes the upper envelope.

The outer edge of the launch cone is at i_0 and this sets a limit on the envelope at

$$r_1 \equiv \frac{v_0^2}{g} \tan i_0, \quad z_1 = \frac{v_0^2}{2g} (1 - \tan^2 i_0) - z_0 \quad (4)$$

where the subscript 1 denotes this ring.

The curve for $i = i_0$ will act as a lower bound for one of the two solutions for $\cot i$ (lower sign) out to (r_1, z_1) :

$$z_m = -z_0 + r \cot i_0 - \frac{gr^2}{2v_0^2} \operatorname{cosec}^2 i_0 \quad (5)$$

where the subscript m denotes this lower bound for populated double solutions for i . This lower bound should provide an internal surface within the plume. Beyond and below (r_1, z_1) the surface equation (5) becomes an outer bound on the plume terminating at $(r_m, 0)$:

$$r_m = \frac{v_0^2}{g} \sin i_0 \left[\cos i_0 + \left(\cos^2 i_0 - \frac{2gz_0}{v_0^2} \right)^{1/2} \right] \quad (6)$$

Should z_1 lie below Io's surface, the surface, equation (3), would terminate at $(r_M, 0)$:

$$r_M = \frac{v_0^2}{g} \left(1 - \frac{2gz_0}{v_0^2} \right)^{1/2} \quad (7)$$

The surface, equation (5), also has an inner intersection with Io's surface at the radius r_p , of the orifice:

$$r_p = \frac{v_0^2}{g} \sin i_0 \left[\cos i_0 - \left(\cos^2 i_0 - \frac{2gz_0}{v_0^2} \right)^{1/2} \right] \quad (8)$$

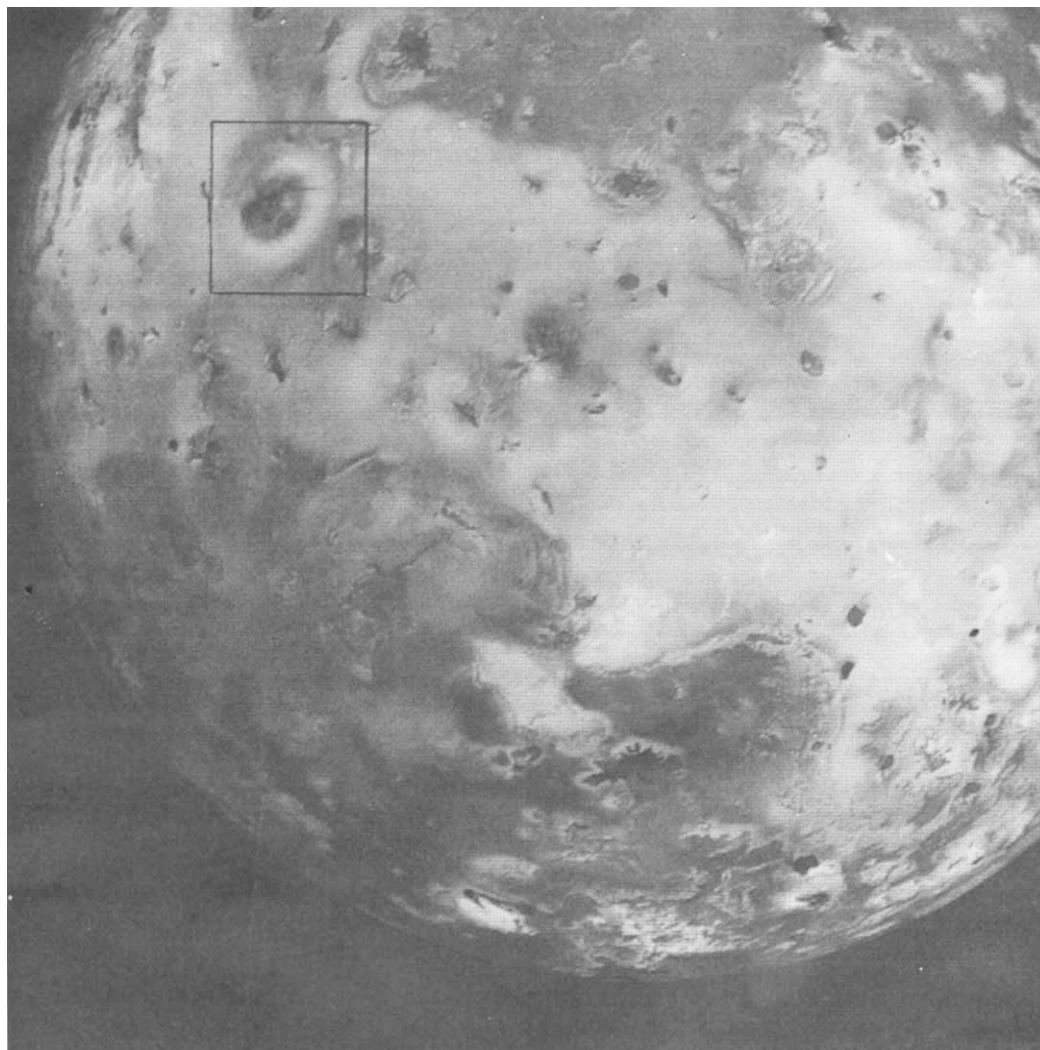


Fig. 1 Box shows Plume 3 at onset of activity as seen from Voyager 1 at 1979 March 4 d 22 h 38 min 12 s UT (FDS 16372.360).

This is the radius of the base of the stalk of the plume.

Measurements of images obtained by the imaging system of Voyager 1 yield the outer radius, r_M or r_m , the radius of the stalk of the plume, r_p , and the height of the plume, h_p (z_M at $r = 0$):

$$h_p = \frac{v_0^2}{2g} - z_0 = \frac{v_p^2}{2g} \quad (9)$$

where v_p is the velocity at departure from the vent. Equations (7), (8) and (9) have the solution

$$\tan i_0 = \frac{2h_p}{r_m + r_p} - \left[\frac{4h_p^2}{(r_m + r_p)^2} - \left(\frac{r_m - r_p}{r_m + r_p} \right)^2 \right]^{1/2},$$

$$\frac{v_0^2}{g} = (r_m + r_p) \operatorname{cosec} 2i_0 \quad \text{and}$$

$$z_0 = \frac{v_0^2}{2g} - h_p \quad (10)$$

There are two plumes which appear to be overwhelmingly from one small vent in an almost horizontal orientation. They are Plumes 1 and 3 discussed by Strom *et al.*² Table 1 presents the results of fitting these measurements on these plumes and includes a comparison of the observed transition point (r_1 , z_1) with the predicted point for Plume 3. Note the blatant failure of our model to fit Plume 3 with r_1 , z_1 (see Plate 5, p. 786) at the outer edges at ground level.

A fundamental difficulty with the ballistic model is that we need an additional force to turn the trajectories of the particles and we are thus led to consider an aerodynamic model.

Jet and shock disk or aerodynamic model

We now consider as an alternative that the combination of particle sizes and rate of gas flow is such that the particles are entrained and transported by the gas into the central portion of the top of the plume with the particles being released into ballistic trajectories after passage through the first part of their descent.

Expansion from a vent at sonic speed cannot produce a Prandtl-Meyer expansion (ref. 3, p. 281–283 and Fig. 4.22, p. 282) into a vacuum unless every molecule freezes or condenses on contact with the surface. Any flow reflected from the nearby surface will produce a shock front near the edge of the stalk of the plume which will initially constrain the horizontal expansion (ref. 4, p. 420–422 and Fig. 203, p. 421). In the absence of a sufficient ambient atmosphere this weak recompression shock will disappear with increasing height and allow a free expansion into a cone at an asymptotic Mach angle:

$$i_0 = \arctan (1/M_0) \quad (11)$$

where M_0 is the corresponding Mach number (see Fig. 3).

The very rapid decompression indicated by the above discussion should leave a very cold gas with negligible thermal energy travelling at an initial velocity, q_p . This leads us to treat an elementary parcel of gas as following a ballistic trajectory until recompression sets in. At that point we expect the presence of a shock front and shall discuss the effects of the presence of such a front.

Consider an element of a vertical cylinder of radius, r , azimuthal angular extent $d\phi$ and vertical extent dz . For this we use

Table 2 Fits to Plumes 1 and 3 for the aerodynamic model

Plume no.	Observed	From (21), (22)*		From (18)	
	i_0 (deg)	q_0^2/g (km)	z_0 (km)	r_m (km)	z_m (km)
1	30	550	29	449	65
3	45	137	7	112	17

Plume no.	From (19)		From (23)		q_p (km s ⁻¹)
	r_m (km)	z_m (km)	h_p (km)	q_m (km s ⁻¹)	
1	212	145	215	0.33	1.00
3	91	39	54	0.16	0.51

* Equation nos given in parentheses.

Comparison and observations

Solution of equations (21) and (22) yields the results in Table 2. The quantity q_m is the velocity just below the shock on the axis of the plume computed by application of the conservation of energy which yields:

$$q = [q_0^2 - 2g(z_0 + z)]^{1/2} \quad (24)$$

The disagreement between the computed height of the plume from the measured r_m , r_p and the measured height probably occurs because the actual trajectories beyond the shock disk are lower than the ballistic trajectories used. This is also indicated by the outer bend in the filaments appearing in Fig. 1. The inner bend is, of course, imposed by the shock front.

The thermodynamic diagram for sulphur dioxide prepared by Smith *et al.*⁵ and their discussion make it clear that we are in a regime in which both the solid and the vapour are present in

roughly equal amounts. In that case passage through the shock front will take place at a ratio of specific heats, close to unity with a large density increase, significant sublimation of entrained crystals of sulphur dioxide and that the temperature immediately above the front will be much lower than that which would occur for a perfect gas. The downward flow would involve expansion into a volume only a few times larger and recompression on reaching the surface could produce another shock front in the shape of a ring with lateral escape, unless most of the gas immediately crystallises onto the surface. The observations do not show such a front so it is either very close to the surface or does not exist owing to rapid crystallisation onto the surface.

Comparison of Plate 7b, p. 789 with Fig. 1 strongly suggests that we are witnessing variable rates of injection of another substance (sulphur droplets?) into a steady gaseous flow (sulphur dioxide).

These results were made possible by all those who have worked on the Voyager Project. Many of these have been individually acknowledged in ref. 1. We thank C. C. Avis, G. W. Garneau, P. L. Jepsen, J. J. Lorre, J. A. Mosher, G. M. Yagi, J. T. Harwood, L. Pieri, and C. Reslock of JPL and also S. A. Collins as experiment representative at JPL. This paper reports one phase of research supported by NASA and JPL under contract NAS7-100 and of research supported by JPL at the Smithsonian Astrophysical Observatory under that NASA contract.

Received 19 July; accepted 3 August 1979.

1. Smith, B. A. *et al. Science* **204**, 951-972 (1979).
2. Strom, R. G., Terrile, R. J., Masursky, H. & Hansen, C. *Nature* **280**, 733-736 (1979).
3. Prandtl, L. *Essentials of Fluid Dynamics* (Hafner, New York, 1952).
4. Oswatitsch, K. *Gas Dynamics* (Transl. Kuerti, G.) (Academic Press, New York, 1956).
5. Smith, B. A., Shoemaker, E. M., Keiffer, S. W. & Cook, A. F. *Nature* **280**, 738-743 (1979).

Volcanic resurfacing rates and implications for volatiles on Io

T. V. Johnson*, A. F. Cook II†, C. Sagan‡ & L. A. Soderblom§

* Jet Propulsion Laboratory, Pasadena, California 91103

† Smithsonian Astrophysical Observatory, Cambridge, Massachusetts 02138

‡ Cornell University, Ithaca, New York 14853

§ US Geological Survey, Flagstaff, Arizona 86001

Resurfacing rates and surface ages on Io are estimated, together with the material ejection and deposition rates of the active volcanic plumes.

THE discovery of current volcanism on Io^{1,2} raises important new questions about the history of volatile material on Io, the rates of material deposition, the age of the currently observed surface, the loss rates of volatiles, and the ultimate fate of volatiles. This article presents an initial comparison of resurfacing rates based on two separate observations and implications for supply of material to the jovian torus. In estimating the resurfacing rates and surface ages we first examine the lack of recognisable impact craters down to the limiting resolution of about 600 m. Second, we estimate from imaging data the material ejection and deposition rates of the active volcanic plumes.

Removal of impact craters

To estimate the age of Io's landforms and hence their removal rate based on the absence of impact craters we must assume some flux of impacting bodies. Although such an estimate has large error bounds we are asking order-of-magnitude questions:

is the resurfacing rate mm per yr, mm per millenium, or mm per aeon? Are the surface landforms hundreds of thousands or hundreds of millions of years old? Two methods have been used to estimate current relative impact rates for the terrestrial planets—Mercury, the Moon and Mars. The first models the distribution and impact rates from telescopically observed asteroids, comets and meteor showers^{3,4}. The second compares the variations in the number of small impact craters (a few kilometres diameter) formed on smooth plains on these bodies following the extremely rapid decline in impact flux ~4,000 Myr ago⁵. Both these techniques suggest that the impact fluxes in the Solar System over a heliocentric range of 0.4 to ~1.5 AU have been roughly uniform, within a factor of two, during the past 10⁹ years or so. The extrapolation to the jovian system suggests that fluxes there cannot be different by more than an order of magnitude from those in the inner Solar System without violating the apparently uniform character of the inner fluxes. Hence for our order of magnitude estimates we will use fluxes quoted for the Moon and Mars⁵ realising that they might be in error by as much as a factor of 10 due to undiscovered populations, Jupiter gravity effects, and so on.

Figure 1 shows the three steps used to estimate that resurfacing rate which will just erase the expected impact craters.

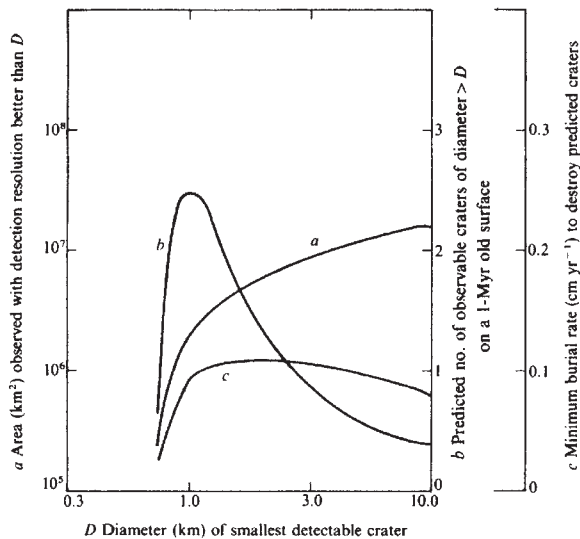


Fig. 1 Estimate of resurfacing rate to destroy impact craters. Shown as a function of crater diameter are: *a*, the areal coverage (in km²) of Io obtained by Voyager 1 on which a crater of diameter *D* or larger could be detected; *b*, the number of craters larger than *D* that would have been seen in the Voyager 1 picture collection if Io's surface were 1 Myr old, assuming a lunar cratering size frequency spectrum and rate; *c*, the minimum rate of burial (or erosion) required to remove the craters so that none are now visible.

Different estimates can be made for different crater sizes. For instance, although larger craters are less frequent, and on the average older, they may take longer to destroy. Furthermore although smaller craters are more frequent they require higher resolution pictures to be recognised. Since in general we have less coverage at the higher resolutions, there is a higher probability of seeing a larger crater than a small one in many cases. Figure 1 shows that for the Voyager 1 pictures the probability of observing a 1 km crater is higher than that for a 600 m crater, although the latter would be far more abundant on the surface. The minimum resurfacing values in Fig. 1 are calculated assuming that the average crater diameter larger than *D* (the diameter of Io) has a diameter $\sim D$, that the craters have a depths of $1/5D$ and that burial by a blanket twice their depth is required to obscure them. The best estimates come from observations of 1–5 km craters, which should be statistically common and easily observed. Our value of the required resurfacing rate 10^{-1} cm yr⁻¹ and should be reliable within a factor of 10. Hence our lower limit is 10^{-2} cm yr⁻¹. Also, resurfacing can be performed by a variety of processes including deposition by volcanic surface flows⁶, volcanic plumes⁷ and surface erosion⁸. Do the observed active plumes, which may be the most active of these processes, provide material at a rate consistent with these burial rates?

Depositional rates of volcanic plume material

To estimate the resurfacing rate from observations of current volcanic activity we start with the dimensions of plumes of volcanic ejecta (r_p is the radius of plume; h_p the height of plume: see Table 1 in ref. 7). The visibility of the plumes in Voyager 1 images implies an optical depth of the order of at least unity in the denser parts of the plume along an optical path of scale $\sqrt{2}r_p$. The plumes are dark when seen in projection against the disk and scatter more strongly at shorter wavelengths when seen off the disk². Accordingly we assume the particles to be small, so their cross section for extinction, σ_p , can be written as

$$\sigma_p = \frac{2\pi^2 K}{\lambda} a^3 \quad (1)$$

where a is the root mean cube of the particle radii, λ the wavelength and K a constant of order unity⁹. Thus, we are considering extinction due to particles with $a < \lambda$, but are ignoring contributions from large particles and Rayleigh scattering

$a \ll \lambda$. For an idealised cylindrical volume circumscribing the plume, we then take the optical depth, τ , to be

$$\tau = \sqrt{2}r_p \sigma_p n_p = \frac{2\sqrt{2}\pi^2 K}{\lambda} a^3 r_p n_p \quad (2)$$

where n_p is the number density of particles.

If we take f_p to be the volume fraction of the cylindrical volume actually filled by plume structure, the total number of particles in the plume, N_p , is

$$N_p = f_p \pi r_p^2 n_p h_p \quad (3)$$

and the total mass of particles, M_p ,

$$M_p = m_p N_p = \frac{4}{3} \pi a^3 \rho_p N_p \quad (4)$$

where ρ_p is the individual density of the particles.

Solving equation (2) for n_p and substituting in equations (3) and (4) we find that:

$$M_p = \frac{\sqrt{2}f_p \tau}{3K} \rho_p \lambda r_p h_p \quad (5)$$

which is independent of particle size a .

The resurfacing rate due to n plumes, R (cm s⁻¹), can then be written in terms of M_p and the time of flight of the particles, t_p . If $R = \sum_{p=1}^n R_p$, then:

$$R = \frac{\sum_{p=1}^n (M_p/t_p)}{4\pi R^2 \rho_p} = \frac{\tau}{6\pi K} \frac{g^{1/2} \lambda}{D^2} \sum_{p=1}^n f_p h_p^{1/2} r_p \quad (6)$$

where D is the diameter of Io and we have used the ballistic flight time given by $t_p = 2\sqrt{2}h_p^{1/2}g^{-1/2}$, g being the acceleration of gravity on Io (~ 180 cm s⁻²). Gas density in the densest parts of the plumes might be enough for drag forces to be important but in the thin parts near the top from which the particles fall back to the surface the ballistic, free fall velocities are supersonic with respect to Io's tenuous atmosphere and the ballistic flight time should be close to the actual flight times of the particles. For example, the ballistic time is about 10^3 s from 100 km altitude.

Taking $\tau \geq 1$, $\rho_p \geq 2$ g cm⁻³, $K \sim 1$, and $\lambda = 4.8 \times 10^{-5}$ cm, Table 1 gives lower limits for the quantity M_p/f_p for each plume, the total for all plumes and R_p/f_p . Using the same quantities in equation (6) yields a value for R/f , if we assume $f_p = f$ for all plumes.

$$\frac{R}{f} \geq 1.1 \times 10^{-10} \text{ cm s}^{-1}$$

The value for R depends on the estimate of f , which must await closer geometric analysis of the images. Assuming that 10% or more of the volume is filled by plume structure we obtain

$$R \geq \sim 3.5 \times 10^{-4} \text{ cm yr}^{-1}$$

This lower limit on the rate is a factor of 100 lower than the lower limit derived from the cratering argument earlier. Our assumptions have been extremely conservative, however. In particular, the derived plume resurfacing rates refer only to a limited size range of particles in plumes actually observed by Voyager 1; larger particles (up to millimetre sizes) could be a significant portion of the mass delivered to the surface. We have also neglected resurfacing due to condensation of the gaseous component of the eruptions, which could at least be equivalent to the particulate mass¹⁰. Also, considerable resurfacing must be done by volcanic flow activity, for which there is abundant evidence in the images^{2,6}. Finally, there are many suggestions in the imaging data of numerous smaller gas/particulate venting below the level detectable as plumes. This low level activity, if constant, could be one of the most important contributors to resurfacing. Taking these factors into account, we conclude that the resurfacing rates required by the crater arguments are in general agreement with the observed level of volcanic activity, but that resurfacing mechanisms in addition to particulate

($a < \lambda$) deposition by the eight known volcanoes are probably necessary unless recent cratering rates at Io have been 100–1,000 times lower than current inner Solar System rates (which we consider unlikely). We will use $10^{-1} \text{ cm yr}^{-1}$ as our nominal rate and $3 \times 10^{-4} \text{ cm yr}^{-1}$ as an extreme lower limit.

Implications for volatile history

The degree of resurfacing described above has profound implications for the geochemistry and geophysics of Io. It implies that large portions of the crust or interior of Io have been turned over or recycled many times in geologic time. We will now investigate the implications of this rapid resurfacing for volatile production rates and volatile loss from Io.

Knowledge of the volatile composition around Io comes primarily from ground-based studies of neutral and ionic species around Io, in the magnetosphere inside Io's orbit, and from Voyager 1 remote sensing and *in situ* analyses of the plasma torus. Materials identified in the torus include (refs in parentheses), Na I (11), Na nuclei (12), K I (13), S II (14), S III (15), S IV (15), S nuclei (12), O II (15–17), O III (10), O nuclei (12), H (18) and possibly S_2^+ or SO_2^+ (17) (see ref. 19 for a summary table). There are obviously still large gaps in our knowledge of the material in Io's vicinity and the inner magnetosphere. However, it seems clear that Io is the most probable source of most, if not all, of this material. The bulk of the data indicates that S and O are the most abundant species, Na less abundant but still significant and other species present only in trace amounts (leaving aside the issue of hydrogen). Thus, despite possible fractionation effects and observational selection factors, we should look for these elements in the material currently escaping from Io. We can estimate escape rates for some of these species assuming Io is the source. For Na, a rate of $\sim 10^8 \text{ cm}^{-2} \text{ s}^{-1}$ is required to produce the observed cloud^{20,21}. Hydrogen was estimated at $10^{10} \text{ cm}^{-2} \text{ s}^{-1}$ (ref. 18) from Pioneer 10 data and the lack of Ly α emission in the Io torus during the Voyager 1 encounter¹⁵ suggests this is an upper limit at present. Injection of S and O ions into the magnetosphere supplies energy which, if radiated away by the observed UV emission features requires a supply rate of $\sim 10^{10} \text{ cm}^{-2} \text{ s}^{-1}$ for these ions¹⁵.

Sulphur has long been regarded as a likely candidate surface material for Io, primarily to explain the very steep decrease in Io's reflectance at short visible wavelengths^{22–24}. The volcanism provides an obvious source for sulphur and sulphur compounds, since these are some of the more abundant materials produced in terrestrial volcanism^{25–27}. Sulphur flows are proposed as one source of coloured features on Io's surface²⁸. Sulphur compounds, particularly S and SO_2 , have been suggested as the driving gas for the explosive volcanisms^{2,10} and SO_2 has been identified in absorption in infrared spectra from the IRIS experiment on Voyager 1²⁹. What is seen in the torus is thus in general agreement with the characteristics of Io and its volcanism (see also ref 19).

We have no direct evidence concerning the water content in Io's volcanism. Water is, of course, a common constituent in terrestrial volcanism (mostly recycled groundwater), and Io may well have had a significant water content at formation^{30,31}. However, based on the absence of any indication of the strong IR H_2O absorptions near $3 \mu\text{m}$, Pollack *et al.*³² place a limit of 10% fractional coverage of water ice on Io and 10^{-3} relative abundance of bound water. Since the areas covered by individual plume ejecta range up to $\sim 10\%$ of the observed disk of Io, water condensation over the area affected by the volcanoes should have been observed; it seems that water must be a minor constituent of Io's volcanism or absent altogether.

We can now take the estimated resurfacing rates and the observed supply rates to the torus to draw some conclusions about the history of volatile supply on Io. Assuming an average of 20 AMU for the species involved, resurfacing rates of $3 \times 10^{-4} \text{ cm yr}^{-1}$ and $10^{-1} \text{ cm yr}^{-1}$ supply $\sim 10^{12}$ atoms $\text{cm}^{-2} \text{ s}^{-1}$ and $\sim 3 \times 10^{14}$ atoms $\text{cm}^{-2} \text{ s}^{-1}$ respectively. It is clear that the current

Table 1 Plume dimensions and resurfacing parameters

Plume no.	r_p^* (km)	h_p^* (km)	M_p/f_p ($\times 10^9 \text{ g}$)	R_p/f_p ($\times 10^{-12} \text{ cm s}^{-1}$)
1	500	280	≥ 63	≥ 66.1
2	105	100	≥ 4.7	≥ 8.3
3	125	70	≥ 4.0	≥ 8.27
4	37	95	≥ 1.6	≥ 2.85
5	100	80	≥ 3.6	≥ 7.07
6	125	80	≥ 4.5	≥ 8.84
7	90	120	≥ 4.9	≥ 7.79
8	70	70	≥ 2.2	> 4.63
Total			≥ 89	113

* From ref. 6.

escape rate of $\sim 10^{10} \text{ cm}^{-2} \text{ s}^{-1}$ can remove only a fraction of the material being brought to the surface by volcanism.

Io has only a very tenuous atmosphere. Limits of surface pressure range from 10^{-6} bar (stellar occultation) to 10^{-8} to 10^{-11} bar from ionospheric models^{33,35}. If only 1% of the material in Io's volcanoes were volatile, then gas would be being supplied at rates of 1 to > 200 times greater than the escape rate. Thus, to avoid building up an atmosphere beyond the observed limits, we must conclude that either there are few or no volatiles associated with Io's volcanoes or that virtually all of the volatiles produced are being recycled back into the surface. An extremely low volatile content is unlikely since some gas is needed to drive the observed eruptions². Thus we conclude that whatever gases are being produced are probably being condensed or otherwise reincorporated onto the surface and only minor amounts are escaping from the volcanoes directly or from a tenuous atmosphere.

The observation of SO_2 gas over a volcanic plume²⁹ and the possibility of an atmosphere containing SO_2 in equilibrium with solid SO_2 at some temperature^{19,29} are consistent with the above arguments. In addition condensed volcanic gases explain some of the puzzling features of Io's spectrum^{32,36}. Hapke³⁷ has suggested, based on transmission spectra of various volcanic materials, that H_2S , SO_2 and possibly S_2O might explain some of the features in Io's reflection spectrum. Fanale *et al.*³⁸ and Smythe *et al.*³⁹ have demonstrated that diffuse reflection spectra of SO_2 frost explain the absorption near $4.0 \mu\text{m}$ very well.

Some of the albedo features on Io's surface may also be understood in terms of condensation phenomena. Inspection of Fig. 12, 14a, 16, 17a and 18 in ref. 2 shows that the deposits of small particles or ash is dark and highly localised. We may match this process with the more localised bright rings outside the inner cores of volcanically active areas (Fig. 18a and b in ref 2). We propose that the bright rings are formed on the outer edges of hot volcanic areas where the surface temperature is low enough to allow sulphur and/or SO_2 frost to deposit from the volcanic gas with the outer limit being set by the outer limit of higher vapour density associated with the plume. An interesting example is the dark area south of Plume 2 shown in Plate 1, p. 782 see Strom *et al.*⁷. The bright deposit from this plume does not cross this dark 'lava lake.' We suggest that this is due to high temperature of this feature⁴⁰ preventing condensation of SO_2 .

An important aspect of Io's volcanism is that it has probably been going on for $(4\text{--}4.5) \times 10^9 \text{ yr}$ (ref. 41). The above range of inferred resurfacing rates implies the supply of anywhere from ~ 0.1 to ~ 10 times the total mass of Io over geological time. This is another way of viewing the conclusion reached earlier that significant amounts of Io's crust and interior have been recycled several times. The current escape rates, on the other hand, only amount to a loss of 10^{-4} of the mass of Io over geologic time. Thus the currently observed rates do not imply significant loss of material from Io since formation. The presence of abundant heavier volatiles such as sulphur is not therefore particularly surprising since only 1% of Io's total sulphur content (if near solar abundance) would have been lost at these rates. However, the ease with which escape occurs may be a strong function of the volatiles involved, and escape rates could have been very

different earlier in Io's history. In fact it seems likely that the large volumes of material brought up by the volcanic activity must have originally contained larger amounts of volatiles such as H_2S and H_2O which were lost through dissociation, ionisation and H escape. This would explain the apparent absence of H_2O combined with active current degassing of sulphur compounds (see also ref. 19).

Implications for supply to plasma and neutral torus

The observation of volcanic eruption plumes on Io also raises the question of what processes are responsible for injecting material into the neutral clouds and plasma torus associated with Io. An initial question is whether material is directly escaping in the eruption process. The height of the observed plumes suggests ejection velocities of $\sim 1 \text{ km s}^{-1}$, below the 2.5 km s^{-1} escape velocity. Since particulate material in an eruption is entrained with the gas, particle velocities greater than the gas velocity are unlikely. Velocities of 1 km s^{-1} are already high for this type of eruptive activity, and higher velocities are unlikely¹⁰. Thus it seems that direct ejection of particulate material with velocities greater than the escape velocity is not responsible for supplying material to the magnetosphere.

Atmosphere loss by either thermal or non-thermal mechanisms of gas supplied by the volcanoes is potentially a major source for the observed escape flux. Jeans escape of sulphur from an atmosphere with a surface pressure of 10^{-10} bar, for instance, would supply $\sim 10^{10} \text{ atoms cm}^{-2} \text{ s}^{-1}$ from the critical level of $\sim 500 \text{ km}$ if the exospheric temperature were 1,200 K and sulphur were a major constituent at that level. Note that the volcanic eruptions may supply material to altitudes near to or even above the exobase. Thus the volcanic activity may bring gases to a level where they will be exposed to escape by thermal or non-thermal mechanisms.

Non-thermal escape mechanisms such as magnetospheric sweeping of ions or dissociative recombination reactions have been suggested as strong possibilities in Io's environment⁴²⁻⁴⁴. These may dominate thermal escape, and here again the volcanic eruptions may play a major role. If Io is 'protected' by the interaction of the co-rotating plasma with a possible Io magnetic field⁴⁵, then the volcanoes may on some sides of Io supply material above the stand-off altitude, allowing interaction with and sweeping by the magnetosphere. Also, although direct escape of particulates from the volcanoes is unlikely, it is possible that some of the very fine micrometre or submicrometre particles in the volcanic plumes may become charged in the plasma environment and be removed by magnetic sweeping. This is at least potentially an important source of dust in the inner magnetosphere. Direct sputtering (see ref. 42) of dust in the plumes by magnetospheric particles is also possible, although the surface area exposed in plumes is less than Io's surface area.

The processes described above may be sufficient to account for current losses of sulphur and oxygen to the torus and for the escape of hydrogen earlier in Io's history. The problem of sodium escape is more difficult, at least partially because we have considerably more information about the neutral species, whereas we only observe the magnetospheric ions for most other species. Three points particularly deserve comment: the apparent constancy of sodium supply since 1973, the observed velocity distribution of the sodium atoms, and the asymmetry of escape required by the cloud geometry.

A constant supply of sodium by volcanoes is not ruled out by the observed volcanic activity, which certainly suggests a rather high average level of activity. However, considering the short lifetime against ionisation implied by observations (20–50 h, ref. 46), major fluctuations in volcanic activity (by a factor of ≥ 2) should have been observed in the sodium data if the sodium escape is directly linked to the volcanoes.

Hapke³⁷ has suggested that Na (and K) are supplied to the torus by evaporation from hot silicate lavas ($T \sim 1,450 \text{ K}$),

possibly carried to great altitudes in volcanic plumes. This seems unlikely for several reasons. First, no temperature of this magnitude has been observed on Io⁴⁰, and models for explosive gas volcanism start with much more modest temperatures¹⁰. Second, any hot silicate which could be ejected by such a gas eruption would cool considerably during its 10 min trip to the top of a plume. Finally, the large Doppler width of the sodium emission lines ($\sim 5 \text{ km s}^{-1}$ FWHM) and the high velocity asymmetry discovered by Trafton are evidence against thermal escape^{47,48}, even if the volcanic activity supplies sodium or sodium compounds to the atmosphere. Since only neutral sodium is involved, non-thermal mechanisms involving ionisation are not possibilities.

Sputtering of atoms from the surface by impacting magnetospheric ions has been suggested as the sodium supply mechanism by Matson and coworkers⁴². Before Voyager, the primary difficulties with this mechanism were the very low efficiencies for sputtering by protons (although ice may be an exception⁴⁹) which were then thought to be the major positive ions in the magnetosphere, and the possible thermalisation of sputtered atoms in Io's atmosphere, requiring yet another energetic event to supply escape energy. The discovery of large densities of sulphur and oxygen ions in the inner magnetosphere has removed the difficulty with yields. For a density of 10^3 cm^{-3} the flux of ions at Io's position in the co-rotational direction is $\sim 6 \times 10^9 \text{ cm}^{-2} \text{ s}^{-1}$. Whether these ions impact Io directly or are precipitated into it by an Io magnetosphere, the energy of the impact will be comparable to the co-rotational energy, $\sim 500 \text{ eV}$ for sulphur at Io's orbit. Heavy ions in this energy range have substantial sputtering yields of the order of unity or greater⁵⁰. Ions of this energy can also withstand several collisions with atmospheric molecules without losing a significant fraction of their energy. Thus the known magnetospheric ion population is sufficient to supply the required sodium flux even for modest sodium abundances in the surface (1–10%).

The major issue remaining is whether sputtered atoms escape Io or are merely supplied to a thermalising atmosphere for later escape. An SO_2 atmosphere (or any other atmosphere that meets the criteria discussed above for volcanic gases that can be reincorporated in the surface) may be considerable at night^{19,29}. Thus this class of atmospheric model may allow atmospheric escape of gases in the daytime and sputtering of surface material at night. Note that sulphur should also be sputtered in some quantity (along with other surface materials), so that volcanic supply of daytime gases might represent a modulation on a lower level, a more steady supply rate.

The shape of the cloud and the velocity profile asymmetry seem to require asymmetric escape of neutrals from Io, most likely from the inner hemisphere, either the trailing or the leading portion thereof^{51,52}. This asymmetry could in principle arise either from asymmetric source or sink processes. The volcanoes are more or less uniformly distributed in longitude^{2,6}, and the night/day effect suggested above averages out to uniform sputtering over an orbit if no other processes operate. Interaction with the magnetosphere remains the best candidate for producing this asymmetry either by creating preferential regions for ion precipitation or possibly by shielding neutrals from electron ionisation in regions inside an Io 'magnetosphere'⁴⁵. It is possible that a day/night modulation combined with a magnetospheric asymmetry could produce the east-west sodium anomalies reported by Bergstrahl *et al.*⁵³ and Goldberg *et al.*⁵⁴, but more detailed models will be necessary to study these possibilities.

Conclusion

The resurfacing rates implied by observations of active volcanism on Io are sufficient to explain the lack of recognisable impact craters on Io's surface. The implied rates of material supply to Io's surface require major recycling of Io's upper crust and mantle and that volcanic volatiles be condensable under Io's surface conditions, consistent with the observation of SO_2 ²⁹ and models of Io's volcanism¹⁰. Atmospheric escape of a small

fraction of the volcanic gas can supply the S and O in Io's torus and these same species, as energetic heavy ions, may be responsible for the supply by sputtering of high velocity neutral sodium.

We thank our colleagues on the Voyager imaging science team, S. Kumar and J. Pearl of the Voyager IRIS team and R. W.

Carlson and D. L. Matson of JPL for stimulating discussions. We also thank F. P. Fanale (JPL) for helpful criticisms. This article presents results of one phase of research carried out at JPL under NASA contract NAS7-100.

Received 16 July; accepted 3 August 1979.

1. Morabito, L. A. *et al. Science* **204**, 972 (1979).
2. Smith, B. A. *et al. Science* **204**, 951-972 (1979).
3. Shoemaker, E. M. in *Impact and Explosion Cratering*, (eds Roddy, D. J., Pepin, R. O. & Merrill, R. B.) 617 (Pergamon, New York, 1977).
4. Wetherill, G. W. in *Impact and Explosion Cratering*, (eds Roddy, D. J., Pepin, R. O. & Merrill, R. B.) 613 (Pergamon, New York, 1977).
5. Soderblom, L. A. in *Impact and Explosion Cratering*, (eds Roddy, D. J., Pepin, R. O. & Merrill, R. B.) 629 (Pergamon, New York, 1977).
6. Carr, M. H. *et al. Nature* **280**, 729-733 (1979).
7. Strom, R. G., Terrile, R., Masursky, H. & Hansen, C. *Nature* **280**, 733-736 (1979).
8. McCauley, J. F. & Soderblom, L. A. *Nature* **280**, 736-738 (1979).
9. van de Hulst, H. C. *Light Scattering by Small Particles*, 179-183 (Wiley, New York, 1957).
10. Smith, B. A., Shoemaker, E. M. & Cook, A. *Nature* **280**, 738-743 (1979).
11. Brown, R. A. in *Exploration of the Planetary System* (eds Woszczyk, A. & Iwaniszewski, C.) 527 (Reidel, Dordrecht, 1974).
12. Vogt, R. E. *et al. Science* **204**, 1003 (1979).
13. Trafton, L. *Nature* **258**, 790 (1975).
14. Kupo, I., Meckler, Y. & Eviatar, A. *Astrophys. J. Lett.* **205**, L51 (1976).
15. Broadfoot, A. L. *et al. Science* **204**, 979-982 (1979).
16. Pilcher, C. B. & Morgan, J. S. *Science* (in the press).
17. Bridge, H. S. *et al. Science* **204**, 987-991 (1979).
18. Carlson, R. W. & Judge, D. L. *Icarus* **24**, 395 (1975).
19. Kumar, S. *Nature* **280**, 758-760 (1979).
20. Brown, R. A. & Yung, Y. L. in *Jupiter*, (ed. Gehrels, T.) 1102 (University of Arizona Press, 1976).
21. Smyth, W. H. & McElroy, M. B. *Planet. Space Sci.* **25**, 415 (1977).
22. Wamsteker, W., Kroes, R. L. & Fountain, J. A. *Icarus* **23**, 417 (1974).
23. Fanale, F. P., Johnson, T. V. & Matson, D. L. *Science* **186**, 922 (1974).
24. Nash, D. B. & Fanale, F. P. *Icarus* **31**, 40 (1977).
25. Bunsen, M. R. *Ann. Chim. Phys.* 3rd Ser. **38**, 215 (1853).
26. Lacroix, A. *Bull. Soc. Min. Fr.* **30**, 219 (1907).
27. Stoiber, R. E. & Rose, W. I. Jr. *Geochim. cosmochim. Acta* **38**, 495 (1974).
28. Sagan, C. *Nature* **280**, 750-753 (1979).
29. Pearl, J. *et al. Nature* **280**, 755-758 (1979).
30. Fanale, F. P., Johnson, T. V. & Matson, D. L. in *Planetary Satellites* (ed. Burns, J. A.) 379 (University of Arizona Press, 1977).
31. Pollack, J. B. & Reynolds, R. T. *Icarus*, **21**, 248 (1974).
32. Pollack, J. B. *et al. Icarus* **36**, 271 (1978).
33. Smith, B. A. & Smith, S. A. *Icarus*, **17**, 218 (1972).
34. Brown, R. A. & Yung, Y. L. in *Jupiter* (ed. Gehrels, T.) 1102 (University of Arizona Press, 1976).
35. Johnson, T. V., Matson, D. L. & Carlson, R. W. *Geophys. Res. Lett.* **3**, 293 (1976).
36. Cruikshank, D. P., Jones, T. J. & Pilcher, C. B. *Astrophys. J. Lett.* **225**, L89 (1978).
37. Hapke, B. *Geophys. Res. Lett.* (submitted).
38. Fanale, F. P. *et al. Nature* **280**, 761-763 (1979).
39. Smyth, W., Nelson, R. & Nash, D. B. *Nature* **280**, 766 (1979).
40. Hanel, R. *et al. Science* **204**, 972-976 (1979).
41. Peale, S. J., Cassen, P. & Reynolds, R. T. *Science* **203**, 892 (1979).
42. Matson, D. L., Johnson, T. V. & Fanale, F. P. *Astrophys. J. Lett.* **192**, L43 (1974).
43. Johnson, T. V., Matson, D. L. & Carlson, R. W. *Geophys. Res. Lett.* **3**, 293 (1976).
44. Cloutier, P. *et al. Astrophys. Space Sci.* **55**, 93 (1978).
45. Kivelson, M. G., Slavin, J. A. & Southwood, D. J. *Science* (in the press).
46. Carlson, R. W., Matson, D. L. & Johnson, T. V. *Geophys. Res. Lett.* **2**, 469 (1975).
47. Trafton, L. *Astrophys. J. Lett.* **202**, L107 (1975).
48. Carlson, R. W. *et al. Astrophys. J.* **223**, 1082 (1978).
49. Brown, W. L. *et al. Phys. Res. Lett.* **40**, 1027 (1978).
50. Carter, G. & Collig, J. S. *Ion Bombardment of Solids*, Ch. 7, 310-353 (Elsevier, New York, 1968).
51. Carlson, R. W. *et al. Astrophys. J.* **223**, 1082 (1978).
52. Matson, D. L. *et al. Science*, **199**, 531-533 (1978).
53. Bergstralh, J. T., Matson, D. L. & Johnson, T. V. *Astrophys. J. Lett.* **195**, L131 (1975).
54. Goldberg, B. A. *et al. Bull. Am. astr. Soc.* **10**, 579 (1978); *EOS* **59**, 1122 (1978).

Sulphur flows on Io

Carl Sagan

Laboratory for Planetary Studies, Cornell University, Ithaca, New York 14853

The colour patterns observed on Io indicate extensive surface flows of quenched molten sulphur

Io has the reddest surface in the Solar System, with the possible exception of Amalthea, which may in fact, be coloured by material lost from Io. Io also displays the largest variation of colour with planetocentric longitude of any other object in the Solar System¹. Voyager 1 close-encounter photomosaics have clarified both the coloration of Io and its surface variation (see, for example the colour images in Smith *et al.*²). Although the Voyager 1 photometric fidelity is not better than 10%, general colour patterns are clear. The surface is mottled with dozens of large, and, probably, thousands of small randomly distributed dark patches, which in most cases are adjacent to red or orange coloured terrain. Equatorial and temperate latitudes are rich in these bright colours. Polar terrain tends to be darker, confirming earlier groundbased impressions. There are also extensive yellow, and white or blue-white, provinces on the satellite's surface. Many of the dark spots are active or recently active shallow volcanic constructs²⁻⁶; seven of the volcanoes have been observed by Voyager 1 in violent eruption. The pattern of albedo features points strongly to a succession of complex effusive and fluvial events. Some dark localised diffuse features exhibit radiating patterns implying downslope flow, such as in Plate 2, p 783. In this image there is a hint of sinuous anastomosing channels, ~300 km long, which are blood-red with some traces of black material down the principal channel. One channel is red proximate to the caldera, fading to yellow distally. In general, dark flows are short and diffuse, red flows longer, and orange and yellow flows tend to be distributed in sheets of considerable area.

A crude similarity between the integrated-disk colour of Io and the spectrum of elemental sulphur led Wamsteker *et al.*⁷ to

suggest the presence of abundant sulphur on the ionian surface. Sulphur has also been identified in the Io-related circumjovian torus both by Earth-based observations⁸ and by Voyager 1 experiments^{9,10}; SO₂ has been identified in the gas phase by Voyager 1 IR spectroscopy¹¹; and it has been suggested that SO₂ frost may explain some of the broadband IR spectral features in the integrated disk of Io^{12,13}. After water and CO₂, sulphur and its compounds are generally the first or second most abundant effluents in terrestrial volcanoes, and fumaroles, in some cases comprising tens of per cent by mass¹⁴. Where fumarole temperatures are $\leq 650^\circ\text{C}$, sulphur gases tend to be the major constituents after steam, and the corresponding volcanic constructs are called solfataras. Sulphur and its compounds have been thought^{2,15} to drive the volcanic plumes because of the absence on Io, unlike the other galilean satellites, of surface water frost in groundbased IR spectroscopic observations. Io has been substantially devolatilised. H₂S is rapidly photodissociated and, because of its very low freezing point, hydrogen has an opportunity to escape efficiently from Io by processes such as Jeans evaporation, or magnetospheric sweeping. As Io becomes more and more oxidising, S_n and SO₂ become prime candidates for the principal volcanic effluents. Ionian effusive flows and sulphur volcanoes may be driven by the lithostatic pressure gradient and explosive decompression of volatiles¹⁶, with the energy derived from tidal dissipation dependent on the resonance lock with Europa¹⁷. However, the following discussion is independent of the mechanics of Io volcanism.

Sulphur coloration

There are perhaps dozens of sulphur allotropes with differing physical, chemical and optical properties derived from differences in both the packing and the bonding of polyatomic sulphur; their melting points vary from 110 to 119°C. Because the melting point of sulphur is ~1,000°C less than that of

silicate, a stochastic distribution of thermal events in the interior of Io should melt sulphur—already known to be abundant near the surface—more often than it should melt silicates. At its melting point liquid sulphur is yellow, intensifying to orange at about 150 °C, red at about 180 °C, and black at about 250 °C (see ref. 18). At the normal boiling point, 444 °C, the composition is mostly cyclic S_8 with some S_6 and S_2 ; at 1,000 °C, sulphur is largely S_2 . Therefore, the species of sulphur in the ionian atmosphere depends strongly on the temperature in the volcanic vent and the rate of cooling. The chemistry of liquid S_n , with $2 \leq n \leq 22$ at least, is complex and incompletely understood^{19,20}.

At 25 °C, solid sulphur shows a steep absorption edge¹⁸ at about 2,850 Å which moves at lower temperatures to shorter wavelengths at about 2.3 Å K^{-1} and at higher temperatures (until the melting point) to longer wavelengths at about 2.0 Å K^{-1} . Thus, at the mean ionian surface temperature, the absorption edge is near 2,500 Å and at the melting point, near 3,000 Å. This absorption edge is partly responsible for the yellow coloration of solid sulphur at room temperature; at liquid nitrogen temperatures orthorhombic sulphur is snow-white because the wings of the absorption edge are sufficiently far into the UV.

Remarkably, if liquid sulphur is rapidly quenched or freeze-dried to low temperatures, its colour is preserved. The faster the freezing, the better the preservation; the higher the initial temperature of the melt, the deeper the hue. These colours have three sources¹⁸: (1) the S_n absorption edge in the violet or near UV; (2) an absorption peak near 4,000 Å due to S_3 ; and (3) an absorption peak at 5,500 Å due to S_4 . S_3 and S_4 are typically present in abundances of ~1% in the frozen melt; they are trapped metastable species and their influence declines markedly at post-quenching temperatures $T \geq 180 \text{ K}$. Therefore, away from calderas and other localised heating, these metastable species are preserved on Io and should affect the coloration of the surface. The colour of quenched liquid sulphur depends on its thermal history and associated metastable allotropes, as well as on impurities.

Terrestrial flows of molten sulphur—such as at the Japanese volcano Siretoko-Iosan²¹—are chocolate brown. However, this is unlikely to be due to a sulphur allotrope; a minimal abundance of organic matter of 1 part in 10^5 can colour sulphur brown¹⁸. The absence of brown coloration on Io would suggest that virtually no organics are present, a conclusion consistent with substantial devolatilisation and escape of hydrogen during Io's history.

The dominant absorption edge of sulphur continues to move to longer wavelengths, through the entire visible spectrum, between the melting point and the normal boiling point, at a rate of approximately 6 Å K^{-1} , or more than twice the rate for the solid¹⁸. Nelson and Hapke²², in a prescient paper, suggested that the integrated-disk coloration of Io is due to quenched liquid sulphur and even proposed, before Voyager, "volcanic fumaroles" as one possible means of melting the sulphur. They suggest that the observed 3300 Å absorption feature on Io is due to quenched liquid sulphur. The above temperature-dependence of the sulphur absorption edge means that a 3,300 Å absorption feature requires liquid sulphur which had been raised to about 155 °C before quenching, that is, temperatures high enough to produce vivid orange-red coloration.

Viscosity, colour and flow time

Figure 1 shows the temperature-dependence of the dynamic viscosity, ν , of liquid sulphur and a rough indication of its colour-dependence. In this model the black patches in and around the volcanic calderas are quenched liquid sulphur originally at $T \geq 250 \text{ °C}$. The connection of such temperatures with volcanic mechanics is discussed in ref. 16. As in terrestrial volcanoes, the incidence of melt in the caldera should be intermittent; there is no Voyager IR evidence for extensive regions with temperatures above the melting point in any one of the few calderas examined to date²³. Because of its comparatively high viscosity (comparable to olivine basalts), black and red-black

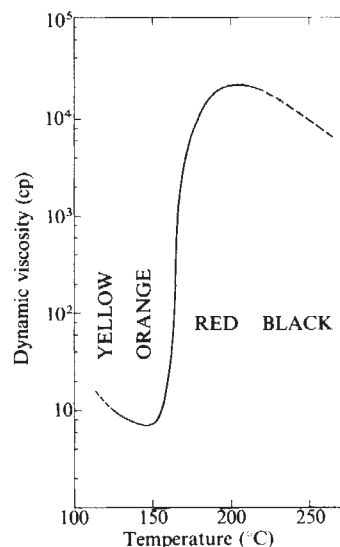


Fig. 1 Viscosity (after the *International Critical Tables*) and colour of molten sulphur as a function of temperature.

liquid sulphur will flow sluggishly and therefore be confined to the vicinity of the caldera. The liquid will cool by radiation and evaporation, the colour will change to red and orange, and the viscosity will drop dramatically, permitting substantial downslope flow and suggesting that the red sinuous deposits are frozen sulphur rivers. When the sulphur melt cools to the orange-yellow and yellow range, just above the melting point, the viscosity is less than 10 times that of liquid water (1 cp at 20 °C), and the melt which is not contained by topographic relief has the opportunity to spread laterally and to form extensive yellow and orange-yellow sheets by flooding the lowland plains. Because of the Stefan-Boltzmann temperature dependence, cooling is about an order of magnitude more rapid in the black and red than in the orange and yellow, and the deposits of the latter coloration should cover more ionian surface area than the former, as is observed. The increase in area covered by the low-viscosity low-temperature melt leads to still faster radiative cooling. Fan-shaped flows, running from red-black interiors to orange-yellow exteriors, are also consistent with the model.

The increase of viscosity with temperature above 157 °C, atypical of most liquids, is produced by the rupturing of the cyclic octatomic form of sulphur to produce long and complex sulphur chains and polymers, with fragmented octagonal forms prominent. At $T > 187 \text{ °C}$ this long-chain structure begins to dissociate thermally. Thus the most interesting and complex sulphur chemistry on Io should occur in the red provinces.

Flow of a sulphur melt in a cylindrical channel on Io will be laminar when the Reynolds number ($V\rho d/\nu$) $< 2,000$, where V is the velocity of flow, ρ the density of liquid sulphur (1.8 g cm^{-3}) and d the channel width. For $d \sim 100 \text{ m}$, and red liquid sulphur, the condition reduces to $V < 10 \text{ cm s}^{-1}$, which may be plausible for high-viscosity flow. Narrower channels are more likely to be laminar. For laminar flow, Poiseuille's equation indicates that the discharge $Q \propto \nu^{-1}$, for a fixed channel diameter and a fixed pressure drop per unit length. The shallow slopes on Io and the lower acceleration due to gravity make both V and Q correspondingly lower than on Earth. A red flow at $V \sim 10 \text{ cm s}^{-1}$ will traverse 100 km in a period of weeks. But for orange and yellow liquid sulphur, ν is so small that the laminar flow condition is unlikely to be met. Thus, black and red liquid sulphur on Io should mobilise slowly and may exhibit laminar flow; while orange and yellow sulphur melts should be correspondingly much faster and fully turbulent. The surface of Io attests to a very complex sequence of overlapping flows. Both the ubiquity of sulphur flows and the likely volcanic mechanisms¹⁶ point to the tapping of an underlying ocean of liquid sulphur, for which the name 'thiosphere' is here proposed.

Consider the cooling of a single eruption of liquid sulphur, sufficient to fill a channel 100 km long, 100 m wide and 10 m

Table 1 Minimum thicknesses d , and rough turnover times t_r , of films of quenched liquid sulphur for $\tau \sim 1$ at $\lambda \approx 5,000 \text{ \AA}$

Colour	Original T	d	t_r
Yellow	$\approx 120 \text{ C}$	tens of centimetres	centuries
Orange	$\approx 140 \text{ C}$	centimetres	decades
Red	$\approx 170 \text{ C}$	millimetres	years
Black	$\geq 250 \text{ C}$	tens of micrometres	days

deep. The sulphur mass is $M \sim 2 \times 10^{14} \text{ g}$ and the radiating surface area is $A \sim 10^{11} \text{ cm}^2$. The cooling time will then be $t_c \gg [(c_p \Delta T + L)M]/[\sigma \bar{T}^4 A]$, where c_p is the specific heat at constant pressure of liquid sulphur $\sim 9.5 \times 10^6 \text{ erg g}^{-1}$, $L \approx 3.8 \times 10^8 \text{ erg g}^{-1}$ is the latent heat of fusion of sulphur, σ is the Stefan-Boltzmann constant, ΔT is taken as $\approx 500 \text{ }^\circ\text{C}$, and the average radiative temperature during the flow $\bar{T}$ is taken as 400 K . The result of this crude calculation is $t_c \gg 2$ months.

Considerably thinner films of liquid sulphur from much more modest eruptions can significantly colour the surface. Table 1 gives rough estimates (data from refs 18 and 24) of the thickness of quenched liquid sulphur necessary to provide significant optical depth in visible light. Because of the high viscosity of hot liquid sulphur, very large deposits should accumulate near the calderas; however, extremely thin quenched films initially at such high temperature would be enough to provide the observed coloration. In contrast, the yellow intercrater plains must be covered by at least tens of centimetres of quenched liquid sulphur.

Table 1 also shows the turnover time t_r for variously coloured deposits to be laid down, assuming a mean planetary deposition rate of volcanic effluvia (other than surface flows) of $10^{-1} \text{ cm yr}^{-1}$ as estimated by Johnson *et al.*¹⁵. Because the deposition rates are not well known¹⁵, and in any case should be much larger than the planetary average near the calderas and somewhat less than the planetary average in the intercrater plains, these figures are very rough. The gross silicate productivity ratio²⁵ of fragmentary ejecta to lava for all volcanic systems on Earth is ~ 5 . If a similar ratio is applied to sulphur on Io, then the mean liquid sulphur flooding rate would be $\approx 10^{-2} \text{ cm yr}^{-1}$, implying (Table 1) that albedo changes in dark circumcaldera regions can take place in periods less than 2 weeks, and changes in red flows in periods less than years. Volcanoes are concentrated at equatorial latitudes⁵, as are the deposits here hypothesised to be recently frozen sulphur melts. Synoptic observations of the surface of Io may uncover large-scale surface albedo and colour changes—particularly in the deeply coloured regions. There are several other possible sources of variable features not involving liquid flows, including the direct deposition of solid volcanic effluvia in the immediate vicinity of calderas and the surface condensation of gas phase effluvia, including SO_2 and S_n . The white deposits on Io might be S_n , rapidly cooled from just above the melting point, but frozen SO_2 is a very promising candidate. The white deposits, like the volcanoes, are also concentrated towards equatorial latitudes.

Decay of sulphur chromophores

A related question concerns the timescale for decay of the sulphur chromophores, particularly the metastable allotropes, at ionian ambient conditions. The colour of quenched liquid sulphur at room temperature persists 'indefinitely' (M. Gouterman, personal communication). Conservatively, we interpret 'indefinitely' as $\gg 1$ month $\sim 3 \times 10^6 \text{ s}$. We adopt exponential colour decay statistics of the form $t = t_0 (\exp(E/kT) - 1)$, where E is the bond energy of the chromophore, and where the equation satisfies the boundary conditions that $t = \infty$ at $T = 0$ and $t = 0$ at $T = \infty$. We then readily find that the characteristic time for chromophore decay at $T = 135 \text{ K}$ greatly exceeds the lifetime of Io for $E > 1 \text{ eV}$. Thus, provided the ionian chromophores are caused by chemical bonds stronger than van der Waals forces or hydrogen bonds (roughly $\sim 0.1 \text{ eV}$), the

coloured deposits on the surface of Io should persist for geologically significant periods of time, from a reaction kinetics standpoint. Charged particle impact and UV radiation damage might, however, change the coloration in shorter times.

Sulphur coverage

Typical ballistic flight times for solids in the ionian volcanic plumes are $\sim 1,000 \text{ s}$. With the minimum mass ejection rates of Johnson *et al.*¹⁵ we calculate productivities for Io volcanoes of 4×10^6 to $3 \times 10^5 \text{ g s}^{-1}$, which is comparable to the Mt Pelee or (1924) Kilauea eruptions, but less than the Surtsey and much less than the Krakatoa eruptions¹⁴. What is striking about Io volcanism is not the individual production rates but the number, frequency, altitude (because of the low atmospheric pressure) and sulphur chemistry of the events. It is of interest to calculate what the mean surface area covered by molten sulphur is at a given moment. Microwave occultation measurements made by Pioneer 10 derive^{8,26}, from the ionisation profile of the nightside ionian atmosphere, a surface pressure of neutral constituents, but for an incorrect composition model. An earlier upper limit, derived from a stellar occultation, of 10^{-6} bar ²⁷ is consistent with the Voyager discovery¹¹ of 10^{-7} bar of SO_2 . We conservatively assume that the atmosphere is composed mainly of S_2 , in vapour pressure equilibrium with a sulphur melt and that the gas phase volcanic effluvia are so hot that they produce significant lateral spreading before they chemically react or freeze out. Rough values for the vapour pressure over black liquid sulphur near $250 \text{ }^\circ\text{C}$ are then $\sim 10 \text{ mbar}$; over red liquid sulphur $\sim 1 \text{ mbar}$, and over yellow $\sim 0.1 \text{ mbar}$. Therefore, in the three cases, for the mean atmospheric pressure to be 10^{-6} bar the fractional coverage of black melts must be $< 10^{-4}$; for red, $< 10^{-3}$; and for yellow $< 10^{-2}$. To reach 10^{-7} bar ($[\text{S}_2] \sim [\text{SO}_2]$) these values must be lowered by one order of magnitude. Thus only a tiny fraction of the surface of Io can be molten at present, although that fraction may be much larger than the corresponding fraction for Earth. Accordingly, the existence of a major flow may be accompanied by a significant increase, at least locally, of surface pressure. Even if 10^{-3} of the surface is at $T \approx 300 \text{ }^\circ\text{C}$ its relative contribution to the net thermal emission from Io will be small.

Vertical relief of several kilometres, tending to localise around calderas on Io, has been recognised⁶. These may be silicate extrusions through the underlying thiosphere. But it is of interest to examine the possibility that they are constructions largely of elemental sulphur which is otherwise in such abundance on the ionian surface. In the Peale *et al.* model¹⁷ of tidal heating in the interior of Io the melting point of sulphur is reached only several kilometres below the surface. Thus, sulphur relief features on Io could have roots extending subsurface only by an amount equal roughly to their height above surface. The tensile and yield strength of such sulphur constructs could be as low as 10^8 dyn cm^{-2} for the density of sulphur and the low acceleration due to gravity on Io and the constructs would still not be crushed by their own weight. (Typical tensile and yield strengths for silicates at room temperature are in excess of 10^9 dyn cm^{-2}). However, the density of solid sulphur exceeds the density of the liquid; if the solid were porous to account for flotation, the strength of the solid would decline markedly, and melting of such a sulphur iceberg at its base would require a rapid production rate for such constructs to maintain the observed steady state population. On the whole, silicate relief on Io may be more likely than sulphur relief (see ref. 4), although surface flows are very probably liquid sulphur and not silicates.

Conclusions

The variegated black, red, orange, yellow and (perhaps) white coloration on Io seems explicable in terms of liquid sulphur, produced in ionian volcanoes and solfataras, and quenched at Io ambient temperatures $\approx 135 \text{ K}$. A number of liquid sulphur flow fronts and frozen rivers appear in the Voyager images. The darker hues may be regions of past sluggish laminar flow, the lighter hues of fast turbulent flow. Opaque layers are at least tens

of micrometres to tens of centimetres thick, depending on colour, and, at present volcanic eruption rates, should have lifetimes against covering by airborne effluvia between days and centuries. The time scale for thermal degradation of chromophores at ambient Io temperatures is significantly longer. Thus coloured liquid flow fronts would have been covered over by solid volcanic effluvia in geologically very short periods of time unless new fluid flows occurred in the same interval. Since the fluvial patterns observed are almost certainly surface rather than atmospheric flows, this suggests that the average renewal time scale of the colour and albedo (but not topographic) features on Io is $\approx 10^3$ yr. In this sense it is the youngest solid surface in the Solar System. The white deposits on Io might be orthorhombic sulphur or SO_2 or other atmospheric effluvia deposited from the atmosphere. However, they must be laid down fast enough not to be covered by flows of molten sulphur. The coloured regions of Io may be where liquid sulphur flows dominate the condensation and deposition of volcanic effluvia; while the white regions represent the opposite case. The dark poles of Io may represent a non-uniform distribution of previous volcanic events, concentrated towards high latitudes. Alternatively, the

lower polar temperatures may produce more rapid quenching of hot sulphur melts, and the preferential preservation of their deep colour. From the deduced neutral atmospheric surface density of Io and the vapour pressure of gas phase sulphur above a melt we provisionally calculate an upper limit to the present fractional surface area of Io which is molten between 10^{-2} and 10^{-5} . Even with the high value for this fraction, the time interval between the appearance of successive major flows of molten sulphur on Io must be less than a decade for the renewal of the entire surface in colour and albedo in $<10^3$ years to be accounted for. But this is consistent with the turnover times calculated independently in Table 1. Thus arguments from volcanic deposition rates and from the vapour pressure over sulphur melts are both consistent with the notion that the surface of Io changes significantly in geologically very short periods of time. Systematic and synoptic observations of Io over several years, as by the Galileo mission, promises the discovery of major surface changes.

I thank B. Smith, M. Gouterman, E. Shoemaker, M. Carr, D. Morrison, D. Pieri, E. Posner and, especially, S. Squyres, for helpful comments. This research was supported by NASA.

Received 18 July; accepted 3 August 1979.

1. Morrison, D. & Burns, J. in *Jupiter* (ed. Gehrels, T.) (University of Arizona Press, 1976).
2. Smith B. *et al. Science* **204**, 951–972 (1979).
3. Morabito, L. A., Synnott, S., Kupferman, P. & Collins, S. A. *Science* **204**, 972 (1979).
4. Carr, M. H., Masursky, H., Strom, R. & Terrile, R. *Nature* **280**, 729–733 (1979).
5. Strom, R. G., Terrile, R., Masursky, H. & Hansen, C. *Nature* **280**, 733–736 (1979).
6. Masursky, H., Schaber, R., Soderblom, L. & Strom, R. *Nature* **280**, 725–729 (1979).
7. Wamsteker, W., Kroes, R. L. & Fountain, J. A. *Icarus* **23**, 417 (1974).
8. Brown, R. & Yung, Y. L. in *Jupiter* (ed. Gehrels, T.) (University of Arizona Press, 1976).
9. Broadfoot, A. L. *et al. Science* **204**, 979–982 (1979).
10. Krimigis, S. *et al. Science* **204**, 998–1003 (1979).
11. Pearl, J. *et al. Nature* **280**, 755–758 (1979).
12. Fanale, F., Brown, R., Cruikshank, D. & Clark, R. *Nature* **280**, 761–763 (1979).
13. Nash, D. & Nelson R. *Nature* **280**, 763–766 (1979).

14. MacDonald, G. A. *Volcanoes*, (Prentice-Hall, New York, 1972).
15. Johnson, T. V., Cook, A., Sagan, C. & Soderblom, L. *Nature* **280**, 746–750 (1979).
16. Smith, B., Shoemaker, E. & Cook, A. *Nature* **280**, 738–743 (1979).
17. Peale, S. J., Cassen, P. & Reynolds, R. T. *Science* **203**, 892–894 (1979).
18. Meyer, B. *et al. Sulfur Research Trends* 53 (American Chemical Society, Washington, 1972).
19. Harris, R. E. *J. phys. Chem.* **74**, 3102 (1970).
20. Cocke, D. L., Abend, G. & Block, J. *J. phys. Chem.* **80**, 524 (1976).
21. Watanabe, T. *Jap. J. Geol. Geog.* **17**, 289 (1940).
22. Nelson, R. & Hapke, B. *Icarus* **36**, 304 (1978).
23. Hanel, R. *et al. Science* **204**, 972–976 (1979).
24. Zanini, M. & Tauc, J. *Appl. Opt.* **15**, 3149 (1976).
25. Rittmann, A. *Volcanoes and Their Activity* 156 (Interscience, New York, 1962).
26. Kliore, A. *et al. Icarus* **24**, 407 (1975).
27. Smith, B. A. & Smith, S. A. *Icarus* **17**, 218 (1972).

Photometric evidence on long-term stability of albedo and colour markings on Io

D. Morrison*, D. Pieri†, J. Veverka‡ & T. V. Johnson†

* Institute for Astronomy, Univ. of Hawaii, Honolulu, Hawaii 96822

† Jet Propulsion Laboratory, Pasadena, California 91103

‡ Laboratory for Planetary Studies, Cornell University, Ithaca, New York 14850

Historical measurements of the albedo and colour of Io are examined for evidence of major resurfacing events.

THE Voyager discovery^{1,2} of intense volcanic activity on Io raises the possibility that endogenic processes could alter the large-scale distribution of colour and albedo on the surface of this satellite in extremely short times on a geologic scale. Estimates of the resurfacing rates on Io from the observed volcanic eruptions and from the lack of observed impact craters have led Johnson *et al.*³ to estimate average resurfacing rates of at least 0.1 mm yr^{-1} , probably $\sim 1 \text{ mm yr}^{-1}$ or more. In a region where an eruption is taking place, the rate could be two or more orders of magnitude larger. Given the size of the observed eruptions⁴, it is not difficult to imagine deposition of an optically thick layer of fresh pyroclastics or recondensed volatiles over tens of thousands of square kilometres on Io in times as short as a few weeks.

Precision measurements of the brightness of Io and the other Galilean satellites have been carried out several times since the pioneering photoelectric work of Stebbins and Jacobsen^{5,6} 50 yr ago; similar measurements of colour extend back to the work of Harris⁷ in about 1950. This article reexamines the historical

measurements of the albedo and colour of Io to determine whether there is evidence of variations, or if not, to set a limit on major resurfacing events resulting in hemispheric brightness changes during the past half century.

Historical data

Extensive broad-band, unfiltered photoelectric measurements of the galilean satellites were first carried out in 1926 and 1927 (refs 5, 6); in the 1950s, less extensive observations on the *UBV* system were undertaken⁷; and by the mid 1970s extensive sets of observations in *UBV* and *uvby* had been obtained by several observers^{8–11}. Because the brightness of the satellites varies with both orbital longitude (rotation) and solar phase angle, many observations are required to separate the functional dependence of brightness on these two parameters. (Alternatively, all of the data necessary to define the mean brightness and rotation curve can be obtained over a very small range in phase, say $5\text{--}7^\circ$, neglecting the phase dependence.) The available observations have been reviewed elsewhere^{10–12}, and the present results are derived from those discussions.

Table 1 summarises the measurements of brightness. In order to minimise variations due to differences among photometric

Table 1 Historical brightness of Io (relative to Ganymede)

Date	Ref.	Mean mag*	Amplitude	Longitude of maximum	Longitude of minimum
1926-27	6	0.24†	0.10‡	100±20	270±20
1951-54	7	0.33	0.18	?	300±30
1973	8	0.43	0.19	95±10	320±10
1973	9	0.41	0.11	100±10	310±15
1974	10	0.42	0.18	100±10	310±10
1974	9	0.40	0.13	100±10	300±15

* [$\bar{V}(\text{Io}, \alpha = 6^\circ) - \bar{V}(\text{Ganymede}, \alpha = 6^\circ)$]

† As transformed by Harris⁷. Since these observations were unfiltered, the transformation introduces an unknown uncertainty of at least 0.1 mag.

‡ The measured amplitude was 0.20; the value here is based on the assumption that Stebbins and Jacobsen^{5,6} measured primarily blue light, and Millis and Thompson⁹ found in 1973-74 that the amplitude in *B* was about 0.1 mag less than the amplitude in *V*.

systems and choice of standards, we have normalised all Io photometry to average values for Ganymede obtained by the same observers during the same time period. We thus assume that Ganymede is unchanging, as indicated by the antiquity of its heavily cratered surface². Because the Stebbins and Jacobsen^{5,6} observations were unfiltered, their data do not transform properly to the standard *V* system. We have listed the mean *V* magnitude as transformed by Harris⁷, but we emphasise the difficulty of achieving a proper photometric transformation of these disparate data to a common system for an object as red as Io. Thus, the absolute brightnesses, even when normalised to Ganymede, remain uncertain to perhaps ± 0.05 mag. In contrast, the amplitude of the light curve should be measured to ± 0.03 mag, and the longitudes of maximum and minimum to $\pm 10^\circ$ (standard deviations).

There is substantial difference in the mean magnitudes listed in Table 1 for 1926-27 and 1951-54, in comparison to more recent data, suggestive of intrinsic changes in albedo. Unfortunately, the changes are not large enough to exclude the possibility that they are due to systematic effects, resulting from problems in the observations or in differing colour response of the instruments. Some observers^{9,10,13,14} have suggested that Io has undergone a monotonic brightening with time; however, we do not see conclusive evidence of a long-term trend of this sort.

The data in Table 1 on light curve amplitude and shape are probably more secure than those on mean magnitude. The amplitude agreement is excellent, except for the 1926-27 data which yield a smaller amplitude in blue than has subsequently been seen, again suggestive of change but insufficient to demonstrate it conclusively. The mean longitude of the minimum, from Table 1, is $\sim 300^\circ$; with a spread from 270° to 320° . The presence of a dark trailing hemisphere on Io is clear, and all data since Stebbins and Jacobsen agree on a deep minimum close to 300° . There is no evidence in these data for any clear temporal trends in the location of dark deposits, and the 50 yr stability of the albedo variations with longitude is large relative to calculated resurfacing times for Io.

A less extensive but significant data set exists for longitudinal variations in colour. Starting with Harris' observations⁷ and continuing through more recent spectrophotometric observations^{8,9,15,16}, all observers have consistently noted that the dark longitude zone near 300° is much redder than the rest of the surface. This general agreement is reinforced by detailed spectral studies over a more limited time span. In particular, medium resolution (4-25 nm) spectrophotometric differences between the bright and dark hemispheres measured by Johnson in 1969¹⁵ and similar measurements reported by Nelson and Hapke¹⁶ during 1975-77 agree within 5% over a range of wavelengths from 300 to 700 nm.

Perhaps the most sensitive way of looking for changes on a surface is by means of polarisation. The degree of polarisation is sensitive not only to the albedo of the surface material, but also to its texture, and can be measured to accuracies of at least 1% for objects such as Io¹⁷. The most extensive set of recent photoelectric polarisation measurements of Io are those of

Zellner and Gradie (personal communication), who find high scatter in some of the Io polarimetry. Although the scatter is difficult to interpret in terms of specific changes on Io, it is probably real, and is evidence of detectable changes on the satellite.

All of the photometric data reviewed here show unusually high scatter for Io. In the past, these apparent variations have been rationalised as being due to difficulties in observing Io in the presence of scattered light from Jupiter. Now, we tend to accept at least part of the variation at face value as indicating major changes affecting areas on the satellite of $\sim 10^4 \text{ km}^2$. However, the overall stability remains striking, with no indication of changes in total albedo over 50 yr $\geq 10\%$, and with the longitudinal asymmetry of albedo and colour preserved to within the accuracy of the measurements.

Interpretation

Is the degree of observed photometric stability of Io consistent with the resurfacing rates deduced from Voyager 1 observations? In order to be detected, changes in colour and albedo must be of an extent and magnitude sufficient to alter the hemispherically averaged optical properties measured by Earth-based photometry. Therefore what are the albedo and colour contrasts typical of volcanic deposits on Io?

The level of contrast in blue light (478 nm) can be judged from Voyager 1 frame FDS 16368.38, centred on longitude 150° (see Plate 7, p. 789). The highest normal reflectivities in this image are 0.76 for extensive white areas that may be SO_2 frost¹⁸. The darkest major regions, which are concentrated near the poles, have a reflectance of 0.20; near the centre of the disk, reflectances range down to about 0.30. Thus, contrast ratios (bright/dark) of factors of two to three are present, with both light and dark material apparently associated with eruptive activity. As shown in Plate 7, these contrasts are present at relatively small spatial scales. Io has a spotty appearance, with spots typically a few hundred kilometres across.

For a contrast ratio of two, a change affecting 2% of the disk, corresponding to a feature with a linear dimension of about 500 km, would produce an integrated brightness change of 0.01 mag. This dimension is characteristic of the largest eruption-associated surface markings seen by Voyager 1; in principle, then, photometry of 0.01 mag precision might detect changes associated with a single very large eruption if it produces material contrasting strongly with the pre-existing surface. However, this level of precision is rarely, if ever, attained in photometry of Io because of scattered light from Jupiter, and the observations shown in Table 1 indicate only that no changes > 0.1 mag have taken place for over 50 yr, and none > 0.03 between 1973 and 1974. The area corresponding to 0.1 mag is 20% of the disk, or $\sim 1,500 \text{ km}$. A history of volcanic eruptions such as those seen by Voyager 1 randomly spaced in longitude would not be expected to result in albedo changes of this magnitude even if the material deposited had the very high contrast level assumed for this estimate.

Colours can be measured photometrically with more precision than albedos, but the historical record on colour measurements is meagre. Examination of blue (478 nm) and orange (590 nm) Voyager images shows that the red areas are 50% more reflective in orange than in blue, while the white areas are essentially neutral. Thus, the covering of a white area by red volcanic deposits, or of red material by white deposits, should produce a measurable (0.02 mag) change in disk-averaged *B-V* colour for areas as small as 700 km on a slide. Careful photometry might thus detect volcanically produced colour changes, but only at the limit of sensitivity of the technique.

There remains the problem, not of detecting changes in the average albedo and colour of Io, but of understanding the presence of major longitudinal differences in the first place. If the surface is constantly renewed by volcanic activity, and if the volcanoes are well distributed in longitude, why should the dark, reddish material be found concentrated in equatorial latitudes

preferentially on the trailing hemisphere over a 50-yr time span? At the resurfacing rates deduced from Voyager 1 of $\sim 1 \text{ mm yr}^{-1}$, an essentially new surface, from the optical point of view, should be generated within a few decades at most. The preservation of the observed longitudinal photometric asymmetry for 50 yr therefore suggests one of the following three explanations: (1) a single eruptive centre (such as P-1), tapping a particular magma source, has continued to dominate the volcanic deposition on the trailing hemisphere for at least 50 yr; (2)

there exists a persistent hemispheric compositional dichotomy in the magma source regions, such that darker material is deposited preferentially on the trailing side in spite of changing individual eruptive centres; (3) the photometric asymmetry is caused by exogenic modification of the surface deposits, for instance, by magnetospheric interactions. We cannot at present distinguish between these three possibilities, which have quite different implications for the nature of volcanism on Io.

Received 16 July; accepted 27 July 1979.

1. Morabito, L. A., Synott, S. P., Kupferman, P. N. & Collins, S. A. *Science* **204**, 972-976 (1979).
2. Smith, B. A. *et al. Science* **204**, 951-971 (1979).
3. Johnson, T. V., Cook, A. F., Sagan, C. & Soderblom, L. A. *Nature* **280**, 746-750 (1979).
4. Strom, R. G., Terrile, R. J., Masursky, H. & Hansen, C. *Nature* **280**, 733-736 (1979).
5. Stebbins, J. *Lick. Obs. Bull.* **13**, 1-11 (1927).
6. Stebbins, J. & Jacobsen, T. S. *Lick Obs. Bull.* **13**, 180-195 (1928).
7. Harris, D. L. in *Planets and Satellites* (eds Kuiper, G. P. & Middlehurst, B. M.) 272-342 (University of Chicago Press, 1961).
8. Morrison, D., Morrison, N. D. & Lazarewicz, A. *Icarus* **23**, 399-416 (1974).
9. Millis, R. L. & Thompson, D. T. *Icarus* **26**, 408-419 (1975).

10. Morrison, D. & Burns, J. A. in *Jupiter* (ed. Gehrels, T.) 991-1034 (University of Arizona Press, 1976).
11. Morrison, D. & Morrison, N. D. in *Planetary Satellites* (ed. Burns, J. A.) 363-378 (University of Arizona Press, 1977).
12. Veverka, J. in *Planetary Satellites* (ed. Burns, J. A.) 171-209 (University of Arizona Press, 1977).
13. Blanco, C. & Catalano, S. *Icarus* **25**, 585-587 (1975).
14. Jones, T. J. & Morrison, D. *Bull. Am. astr. Soc.* **7**, 384 (1975).
15. Johnson, T. V. *Icarus* **14**, 94-111 (1971).
16. Nelson, R. M. & Hapke, B. W. *Icarus* **36**, 304-329 (1978).
17. Veverka, J. in *Planetary Satellites* (ed. Burns, J. A.) 210-231 (University of Arizona Press, 1977).
18. Fanale, F. P., Brown, R. H., Cruikshank, D. P. & Clark, R. N. *Nature* **280**, 761-763 (1979).

Identification of gaseous SO_2 and new upper limits for other gases on Io

J. Pearl*, R. Hanel*, V. Kunde*, W. Maguire*, K. Fox†, S. Gupta‡, C. Ponnamperna† & F. Raulin‡

* Goddard Space Flight Center, Code 693.2, Greenbelt, Maryland 20771

† University of Tennessee, Cumberland Avenue, Knoxville, Tennessee 37916

‡ University of Maryland, College Park, Maryland 20740

Gaseous SO_2 has been identified on Io. The estimated abundance of 0.2 cm atm is consistent with an atmosphere in equilibrium with solid SO_2 at the local surface equilibrium temperature. Preliminary upper limits for several gases, including sulphur compounds and other terrestrial volcanic emissions have been derived. Io is apparently depleted in hydrogen, carbon and nitrogen.

ONE objective of the Voyager IR experiment (IRIS) is to search for and analyse atmospheres on the jovian satellites¹. The possibility of an atmosphere on Io has been studied for some time²⁻⁶, and as a result, upper limits have been established for many gases and for the total surface pressure. The possible existence of various sulphur compounds on the surface^{7,8}, together with the discovery of active volcanism^{9,10} suggest that further investigations should include searches for additional gases which are products of terrestrial volcanism, and sulphur compounds in particular.

Previous spectroscopic investigations of the atmosphere of Io have been restricted to the visible^{2,3} and near IR⁶. Although it contains the rotational and vibrational signatures of many gases, the thermal IR region has not been investigated because of difficulties in obtaining high-quality spectra. The Voyager 1 IR spectroscopy experiment, however, has now made this region available.

During the flyby of Io, over 200 IR spectra of the satellite have been obtained at a spectral resolution of 4.3 cm^{-1} and spatial resolutions down to 50 km (ref. 11). The useful spectral range of these data extends from 180 cm^{-1} to an upper limit which depends on surface temperature: about 650 cm^{-1} at 90 K and $1,100 \text{ cm}^{-1}$ at 130 K. One of the most interesting subsets of the data covers an unexpectedly warm region containing a volcanic feature near which a plume (Plume 2 of Smith *et al.*¹⁰) originates. Spectra of adjacent regions show surface brightness temperatures near 130 K at 250 cm^{-1} , with a slight rising trend with

increasing wavenumbers. However, spectra over the warm region show brightness temperatures in excess of 210 K at large wavenumbers. This effectively extends the useful spectral range of the data by increasing the wavenumber at which the signal-to-noise ratio equals unity to beyond $1,900 \text{ cm}^{-1}$.

Identification of SO_2

Seven spectra over the 'hot spot' were averaged to provide the data for this analysis. The portion of the resulting spectrum between $1,000$ and $1,500 \text{ cm}^{-1}$ is shown in Fig. 1 and an expanded version in Fig. 2. We identify the feature between $1,320$ and $1,380 \text{ cm}^{-1}$ as the ν_3 band of SO_2 . In addition to ν_3 , SO_2 has two other IR-active fundamental bands in the IRIS spectral range: ν_1 at $1,152 \text{ cm}^{-1}$ and ν_2 at 518 cm^{-1} . The strengths of the ν_1 , ν_2 , and ν_3 bands are in the ratio 1:1:8, respectively¹². As a result of this, of lower thermal contrast at lower wavenumbers, and of possible additional radiative transfer effects, only the strongest band is clearly evident in the IRIS spectrum. A faint suggestion of ν_1 seems present in Fig. 1. The ν_2 band lies in a region strongly contaminated by presently unidentified spectral features; however, no clearly identifiable structure assignable to ν_2 is apparent.

To confirm the identification of ν_3 and to obtain an estimate of the sulphur dioxide abundance, a simple model consisting of a single homogeneous layer of SO_2 at 130 K has been used to generate synthetic comparison spectra. Line positions, strengths, and collision halfwidths are from A. Goldman (personal communication). Doppler and Lorentzian broadening are both taken into account. Several abundances, with their corresponding surface pressures, have been considered; in keeping with the homogeneous model, the pressure used in the computations is taken as half of the corresponding surface pressure. A continuum level is chosen which matches that of the data on both sides of the band. Resulting synthetic absorption spectra for SO_2 abundances of 0.1 and 1.0 cm atm (corresponding to surface pressures of 5.2×10^{-8} and 5.2×10^{-7} bar, respectively) are also shown in Fig. 2. The band contour in the data is

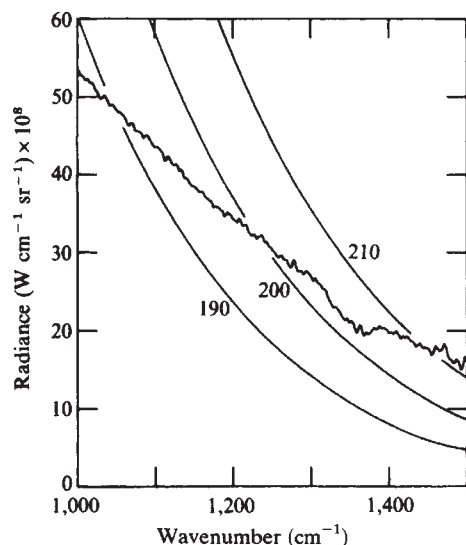


Fig. 1 Portion of Io spectrum over hot spot. The weak ν_1 ($1,150\text{ cm}^{-1}$) and the stronger ν_3 ($1,361\text{ cm}^{-1}$) bands of SO_2 fall in this spectral range. The smooth curves represent black body spectra at the indicated temperatures.

rather well reproduced; a best fit seems to lie between the two calculated curves shown, but closer to the one representing 0.1 cm atm . We adopt 0.2 cm atm as the most likely abundance, with an estimated probable error of a factor of 2. Refinement of this abundance estimate requires a more realistic radiative transfer model.

Upper limits for undetected constituents

Several other gases were searched for in the Io spectrum, but were not detected. Upper limits were calculated using an expression derived from the radiative transfer equation¹:

$$u = \frac{nN}{[B(T_s) - B(\bar{T})]} \times \frac{\Delta\nu}{S(\bar{T})}$$

Here u is the upper limit abundance; n is the margin above the noise level (N) which is considered adequate for detection ($n = 2$ is adopted); $\Delta\nu$ is the instrumental resolution (4.3 cm^{-1}); $[B(T_s) - B(\bar{T})]$ is the thermal contrast between the lower boundary and the atmosphere, where B is the Planck function, T_s is the background brightness temperature, and $\bar{T}$ is the mean atmospheric temperature; $S(\bar{T})$ is the line strength of the atmospheric constituent. Where possible, intervals with strong Q -branches were chosen. The total line strength used was obtained by summing all lines within the appropriate 4.3 cm^{-1} resolution element. For calculation purposes, two mean gas temperatures, 130 and 250 K were chosen; the first is for a gas in equilibrium with the surface surrounding the 'hot spot', and the second is for a warm atmosphere model¹³. The background temperature, T_s , is taken from the measured Io spectrum. Because of the nonisothermal character of the spectrum, evident in Fig. 1, T_s is a function of wavenumber. The temperature dependence of the line strength and the variation of the noise level with wavenumber have also been taken into account.

Preliminary upper limits for 11 gases, including four sulphur compounds (COS , CS_2 , SO_3 , and H_2S), are listed in Table 1. The line strength for COS is taken from Penner¹⁶; that for CS_2 is derived from Robinson¹⁷ and Smith¹⁸. The strength for SO_3 is obtained from Bent and Ladner¹⁹ and from Majkowski *et al.*²⁰, as follows. The latter group measured the absorption coefficient of ν_3 at $1,391\text{ cm}^{-1}$. From this the absorptions of the fundamentals ν_2 at 500 cm^{-1} and ν_4 at 532 cm^{-1} were obtained by scaling the transmission spectra of Bent and Ladner.

The strongest band of H_2S in the spectral range of the data (ν_2) is relatively weak; hence, the upper limit has not been derived from the IRIS spectrum, but from an extension of the analysis of the 1971 occultation of Beta Scorpil by Io. From the geometry of the eclipse and the rapid disappearance of the stellar signal, Bartholdi and Owen² showed that an upper limit to Io's surface

pressure in atmospheres is given by $2 \times 10^{-13} \times (T/M)^{3/2} / \rho$, where T , M , and ρ are the temperature, molecular weight, and refractivity of the gas. Assuming that the gas is in equilibrium with the surface, and noting that the occultation points are near the terminator, we take $T = 110\text{ K}$. The refractivity of H_2S is approximately 6.3×10^{-4} (ref. 21). Thus, an upper limit to the surface pressure is $2 \times 10^{-8}\text{ bar}$, which corresponds to an upper limit abundance of 0.07 cm atm . A better limit for H_2S may be obtained from the IRIS data when line strengths for the pure rotation spectrum become available.

We have generated the line positions and strengths for CO_2 , CH_4 , and NH_3 from standard quantum mechanical formulations using molecular constants based on laboratory data. For O_3 and H_2O , the line positions and strengths of McClatchey *et al.*²² are used.

Discussion

The region of Io observed during these measurements contains three features: Plume 2 (ref. 10), a warm region at $\sim 290\text{ K}$ which covers roughly 10% of the instrument field of view, and a remaining background area at $\sim 130\text{ K}$ (ref. 11). The observed SO_2 could originate from any or all of these sources.

Since sulphur dioxide is generally the major sulphur bearing component of terrestrial volcanic gases²³, an obvious source for the gas is the volcanic plume. The resurfacing rate due to this plume can be used to estimate the amount of gas within it. By assuming that the optically thick (particulate) portion of the plume occupies 10% of the cylinder defined by the plume boundaries, and that the gas-to-solid mass ratio is 2, the data of Table 1 of Johnson *et al.*²⁴ indicate that the plume contains at least 10^9 g of gas. As uncertainties in the resurfacing rate and in the gas-to-solid ratio are unlikely to provide an error of two orders of magnitude, 10^{11} g seems to be a safe upper limit to the gas content of the plume. Over the warm area, which provides the best thermal contrast, the derived SO_2 abundance implies $2 \times 10^{11}\text{ g}$ of SO_2 . Thus, only in the most extreme case would the plume contribution be a significant fraction of the total.

The temperature of the warm area is far above the melting point of SO_2 (198 K); explosive boiling and extensive condensation as clouds would be evident if a substantial amount of liquid SO_2 were present²⁵. Theoretical arguments^{26,27} suggest that a cooling sulphur extrusion is more likely. Outgassing of SO_2 originally dissolved in such a melt will certainly occur; however, once the material has cooled this extensively (95 K below the melting point), such activity should be much reduced due to the solidification of at least the surface layer; further data (see below) suggest that this is the case.

From spectra recorded near to, but not including the hot spot, the surface temperature of the surroundings is determined to be $\sim 130\text{ K}$. At this temperature, accumulation of solid SO_2 by

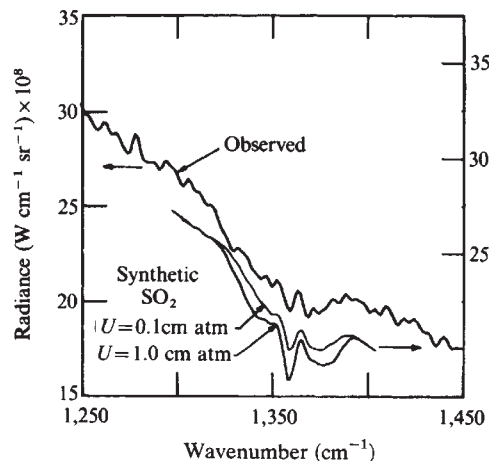


Fig. 2 Comparison of Io and synthesised spectra in the vicinity of the ν_3 band of SO_2 . The synthesised spectra represent absorption by homogeneous layers of 0.1 and 1 cm atm of pure SO_2 gas at 130 K and pressures of 2.6×10^{-8} and $2.6 \times 10^{-7}\text{ bar}$, respectively. The measured and calculated spectra are vertically displaced for clarity.

Table 1 Upper limits of various gases on Io

Gas	Band used (cm ⁻¹)	Vapour pressure abundance (cm atm) above solid at 130 K	Upper limit (cm atm) for	
			$T_{\text{gas}} = 130 \text{ K}$	$T_{\text{gas}} = 250 \text{ K}$
COS	ν_1 859	510*	3.8×10^{-4}	1.5×10^{-4}
CS ₂	ν_3 1,535	1.4×10^{-1} *	3.5×10^{-5}	2.8×10^{-5}
SO ₃	ν_2 497	1.6×10^{-6} *	6.5×10^{-5}	2.4×10^{-5}
H ₂ S	—	56	—	7×10^{-2} †
CO ₂	ν_2 667	1,050	1.5×10^{-4}	5.1×10^{-5}
O ₃	ν_1 1,042	‡	1.9×10^{-3}	1.1×10^{-3}
N ₂ O	ν_1 589	4,780	7.4×10^{-3}	2.2×10^{-3}
H ₂ O	Rot. 254	6.4×10^{-7} *	9.2×10^{-3}	1.9×10^{-4}
CH ₄	ν_4 1,306	‡	1.0×10^{-2}	1.7×10^{-3}
NH ₃	ν_2 931	11	1.4×10^{-3}	3.4×10^{-4}
HCl	Rot. 206	1.8×10^{-4}	3.8×10^{-3}	3.7×10^{-4}

With the exception of H₂S, the analysis is based on the absence of spectral features in the Io spectrum containing a hot spot. The H₂S limit is based on a stellar occultation (see text). For comparison, abundances of gases in equilibrium with their solid phases are included. These are evaluated from ref. 15 unless otherwise indicated.

* Based on data extrapolated from ref. 14.

† From stellar occultation using $T_{\text{gas}} = 110 \text{ K}$ (see text).

‡ Above melting point.

precipitation of solid particles from the plume, by direct condensation of the plume upon contact with the surface, and possibly by extrusion of liquid SO₂ through surface fractures²⁵ may occur. Indeed, the 4.1 μm feature in spectra of Io obtained from the Earth^{28,29} has been identified as the $\nu_1 + \nu_3$ band of solid SO₂^{30,31}. Thus, it seems likely that an SO₂ atmosphere in approximate equilibrium with the vapour pressure of solid SO₂ could become established.

The surface pressure corresponding to our derived abundance of SO₂ is 1.0×10^{-7} bar, with an uncertainty of a factor of 2. Excellent agreement therefore exists with the vapour pressure of SO₂ at 130 K (1.4×10^{-7} bar), as determined by Wagman³². However, our measurement is significantly higher than the 10^{-8} – 10^{-9} bar limit set for the neutral atmospheric pressure from the Pioneer 10 occultation measurements⁴. On the other hand, our observation was made near 13.00 h LT, whereas the Pioneer occultations occurred near 05.30 and 17.30 h LT. For these local times and comparable low latitudes, IRIS measurements indicate surface temperatures below 110 K. Extensive diurnal surface temperature variations, coupled with the strong dependence of SO₂ vapour pressure on temperature, will produce very steep gradients in atmospheric density with local time, and on collapse of the SO₂ atmosphere to an exosphere on the night side of Io; extreme cases of such behaviour, which also show the effects of trapping, are evident in exospheric models for the lunar atmosphere^{33,34}. At 110 K, the vapour pressure of SO₂ is 2.5×10^{-10} bar, which is more compatible with the Pioneer results. It is more difficult to reconcile the present determination with the much lower neutral atmosphere estimates by Johnson, Matson and Carlson⁵.

Recent theoretical work^{24,35} also supports the presence of SO₂ in the volcanic plumes and as a neutral atmosphere; sulphur dioxide seems capable of providing the sulphur and oxygen to the Io torus, and of supporting the measured ionosphere.

A strongly blue component, assigned to Rayleigh scattering by particulates, has been observed in images of the volcanic plumes^{10,24}. Minute particles of condensed SO₂ could account for this. Close inspection of the IR data and the synthetic spectra in Fig. 2 reveals a possible weak absorption between 1,320 and 1,340 cm⁻¹. In the solid state, the ν_3 band of SO₂ is shifted into this region, as is shown in Fig. 3. Comparison of the feature of the solid with the data does not show complete agreement; hence, no identification can yet be made. We are now conducting laboratory studies to provide quantitative information on this band.

Unfortunately, with two exceptions, it has not been possible to map the suggested SO₂ atmosphere over the planet because the surface is generally too cold to provide an adequate signal in the strong ν_3 band. The exceptions are two additional smaller

hot spots observed on the dark side of Io. Based on preliminary pointing information, these occur at (32°, 103°) and (28°, 194°). Taking account of the 425 km diameter IRIS field of view at the time of the observations, the first of these can be identified with Plume 5 (ref. 10); to date, no plume has been identified with the second. In both cases, the surface temperature surrounding the hot spot is below 100 K. Although the signal-to-noise ratio in these spectra is poor, no signature of ν_3 is apparently present; this is consistent with the low vapour pressure of 3.9×10^{-12} bar for SO₂ at this temperature. It also suggests that outgassing from the material at 290 K is not at present a major source of SO₂.

The atmosphere will be affected by factors such as local gradients associated with global temperature variations, and loss mechanisms due to ionisation, surface chemical reactions, and burial by the dominant gas constituent. However, if it is assumed that production mechanisms, such as plumes and extrusive activity, have been active long enough for a steady state condition to be established, then if loss processes are not significant, the vapour pressures of volatiles become important in the consideration of upper limits. If the vapour pressure exceeds the pressure corresponding to the derived limits, then the limits represent true upper limits on possible gaseous constituents, and thereby on their solid phases.

The limits for CH₄, NH₃, H₂S and N₂O of our preliminary results in Table 1 are improvements over previous determinations by Fink *et al.*⁶; limits for CO₂ and H₂O are improvements over values obtained by Bartholdi and Owen². Note that the presence of an SO₂ atmosphere does not affect the derivation of the H₂S upper limit, since at the appropriate gas temperature (assumed equal to the surface temperature of ~110 K), the vapour pressure of SO₂ is so low that it will not be detectable; the vapour pressure of H₂S, however, will support a substantial atmosphere. As indicated in Table 1, for all of the gases except H₂O and SO₃, the limits correspond to pressures far below the respective vapour pressures at 130 K. Thus, we conclude that Io is deficient in these materials. The low abundances of carbon and nitrogen containing gases are significant, the implied depletion being consistent with the lack of C and N in the Io torus³⁷.

The vapour pressures of H₂O and SO₃ at 130 K place these gases substantially below our detection threshold, so the present limits only imply that there are no strong non-equilibrium sources of these gases within the instrument field of view. The sensitivity of the IRIS instrument to condensed phases on the satellite surface is much reduced by very low thermal contrast. Thus, the absence of near-IR absorption features of H₂O still provides the best upper limit for this constituent²⁹. Combined with the low abundance of H₂S, however, it seems that Io is also deficient in hydrogen. The low H₂S abundance also conflicts with Hapke's suggestion that this gas should be present on Io³⁸.

Conclusions

The thermal IR spectrum of Io indicates the presence of a sulphur dioxide atmosphere which is in equilibrium with solid SO₂ at the local surface temperature. A likely source of this

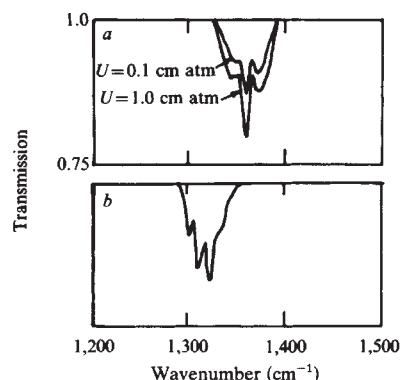


Fig. 3 a, Transmission spectrum of SO₂ gas computed from line-by-line integration. b, Measured SO₂ ice spectrum. Adapted from ref. 36.

material is volcanic activity. Losses to space may account for the sulphur and oxygen observed in the Io torus; condensation and precipitation may account in part for the burial of impact craters. Upper limits for several other gases have been set, which indicate a depletion of hydrogen, carbon, and nitrogen in the satellite; this is consistent with the absence of these species in the torus. All of these results provide strong support for major sulphur volcanism^{26,27} driven by tidal interactions³⁹.

Received 16 July; accepted 26 July 1979.

1. Hanel, R. *et al. Space Sci. Rev.* **21**, 129 (1977).
2. Bartholdi, P. & Owen, F. *Astr. J.* **77**, 60 (1972).
3. Smith, B. A. & Smith, S. A. *Icarus* **17**, 218 (1972).
4. Kliore, A. J. *et al. Icarus* **24**, 407 (1975).
5. Johnson, T. V., Matson, D. L. & Carlson, R. W. *Geophys. Res. Lett.* **3**, 293 (1976).
6. Fink, U., Larsen, H. P. & Gautier, T. N. *Icarus* **27**, 439 (1976).
7. Wamsteker, W., Kroes, R. L. & Fountain, J. A. *Icarus* **23**, 417 (1974).
8. Nash, D. B. & Fanale, F. P. *Icarus* **31**, 40 (1977).
9. Morabito, L. A., Synnott, S. P., Kupferman, P. N. & Collins, S. A. *Science* **204**, 972 (1979).
10. Smith, B. A. *et al. Science* **204**, 951–972 (1979).
11. Hanel, R. *et al. Science* **204**, 972–976 (1979).
12. Tejwani, G., Fox, K. & Corice, R. J. *Chem. phys. Lett.* **18**, 365 (1973).
13. McElroy, M. B. & Yung, Y. L. *Astrophys. J.* **196**, 227 (1975).
14. Hodgman, C. D., Weast, R. C. & Selby, S. M. (eds) *Handbook of Chemistry and Physics* 38th edn, 2143 (CRC, Cleveland, 1956).
15. Washburn, E. W. *International Critical Tables of Numerical Data. Physics, Chemistry, and Technology* Vol. 3, 203 (McGraw-Hill, New York, 1928).
16. Penner, S. S. *Quantitative Molecular Spectroscopy and Gas Emissivities* (Addison-Wesley, Reading, 1959).
17. Robinson, D. Z. *J. Chem. Phys.* **19**, 881 (1951).
18. Smith, D. F., Jr & Overend, J. *Spec. Acta* **26A**, 2269 (1970).
19. Bent, R. & Ladner, W. R. *Spec. Acta* **19**, 931 (1963).
20. Majkowski, R. F., Blint, R. J. & Hill, J. C. *Appl. Opt.* **17**, 975 (1978).
21. Allen, C. W. *Astrophysical Quantities* (Athlone, London, 1955).
22. McClatchey, R. A. *AFCRL Atmospheric Absorption Line Parameters Compilation* (AFCRL-TR-73-0096, 1973).
23. White, D. E. & Waring, G. A. *U.S. Geol. Survey Prof. Paper*, 440-K (1963).
24. Johnson, T. V., Cook, A. F., Sagan, C. & Soderblom, L. A. *Nature* **280**, 746–750 (1979).
25. McCauley, J. F., Soderblom, L. A. & Smith, B. A. *Nature* **280**, 736–738 (1979).
26. Sagan, C. *Nature* **280**, 750–753.
27. Smith, B. A., Shoemaker, E. M. & Cook II, A. F. *Nature* **280**, 738–743 (1979).
28. Cruikshank, D. P., Jones, T. J. & Pilcher, C. B. *Astrophys. J. Lett.* **225**, L89 (1978).
29. Pollack, J. B. *et al. Icarus* **36**, 271 (1978).
30. Fanale, F. P., Brown, R. H., Cruikshank, D. P. & Clark, R. N. *Nature* **280**, 761–763 (1979).
31. Smythe, W. D., Nelson, R. M. & Nash, D. B. *Nature* **280**, 766 (1979).
32. Wagman, D. D. *Data Sheet, Chem. Thermodynamics Data Center, U.S.A.* (National Bureau of Standards, Washington, 1979).
33. Vogel, U. *Planet. Space Sci.* **14**, 1233 (1966).
34. Hodges, R. R. *The Moon* **14**, 139 (1975).
35. Kumar, S. *Nature* **280**, 758–760 (1979).
36. Giguere, P. A. & Falk, M. *Can. J. Chem.* **34** (1956).
37. Broadfoot, A. L. *et al. Science* **204**, 979–982 (1979).
38. Hapke, B. *Geophys. Res. Lett.* (submitted).
39. Peale, S. J., Cassen, P. & Reynolds, R. T. *Science* **203**, 892–896 (1979).

The stability of an SO₂ atmosphere on Io

S. Kumar

Jet Propulsion Laboratory, Pasadena, California 91103

Measurements of Io's volcanism and the plasma torus suggest that SO₂ is a major constituent of its atmosphere.

Recent Voyager 1 measurements on Io's volcanism and the plasma torus suggest that Io could have a thermosphere with SO₂ as the major gas at a surface density of 10¹¹–10¹³ cm⁻³ on the dayside. The bulk of the atmosphere is most likely protected from Jupiter's co-rotating plasma at Io's orbit by either a magnetosphere of Io or by slowing down of the plasma in a comet-like interaction. Yet the scavenging rates above the magnetopause or ionopause are high enough that this atmosphere must be continually replenished by the volcanic activity, otherwise the atmosphere would disappear rapidly. Other minor constituents expected are argon, SO₃, S₂, O, S, and CO₂.

The maintenance of a well developed ionosphere observed on Io^{1,2} has been puzzling due to the relatively strong ram pressure of Jupiter's co-rotating plasma at Io's orbit which would sweep away any gravitationally bound ionosphere³. If the ionosphere were protected, models proposed using NH₃ and Na (refs 3–6) would explain the observed ionospheric densities. The recent Voyager encounter of Io has yielded a large quantity of diversified data (see Table 1) from which an entirely different picture of Io is unfolding and which compels us to revise our thinking about Io's atmosphere and ionosphere.

Active volcanism observed on Io^{7,8} could provide a continuous source of gases from Io's interior. The velocity of the ejected material is ≤ 1 km s⁻¹ (ref. 7), comfortably less than the escape velocity of 2.56 km s⁻¹ on Io. Hence, any direct escape is negligible. The resurfacing rate due to the dust ejected from the volcanoes is at least 10⁻³ m yr⁻¹ (ref. 9). A comparable amount of gas would be needed to drive the dust out of the volcanoes. This leads to a gas outflux of $\sim 10^{14}$ molecules cm⁻² s⁻¹ averaged over the globe. In comparison, a supply rate of only $\sim 10^{10}$ cm⁻² s⁻¹ for ions of S and O is needed into the plasma torus¹⁰ if ion co-rotation energy is the primary source of heating

We thank S. Kumar, R. Samuelson, T. Johnson, B. Smith, and H. Masursky for helpful discussions; also L. Herath, F. Rockwell, J. Frost, and P. Corbin for the calibration, data handling, and computer support. Finally, we thank L. Horn for observational planning.

Note added in proof: Interestingly, many of our conclusions have also been reached by W.-H. Ip and W. I. Axford (*Nature*, submitted).

the torus electrons. An estimate of the total gas supply rate can be made by requiring that the observed maximum plasma density of 4,000 cm⁻³ (ref. 11) be maintained over the diffusion time of ions out of a 2 R_J thick plasma torus (R_J = radius of Jupiter). Eviatar *et al.*¹² have estimated a diffusion time of 20 days to drift 1 R_J in radial sense at Io's orbit. This leads to a supply rate of $\sim 10^{11}$ cm⁻² s⁻¹. Mass loading would, however, reduce the velocity of ions and hence increase the diffusion time. Based on Voyager 1 plasma data which suggest massive mass loading of co-rotating plasma by Io, the estimated supply rate is $\sim 10^{12}$ cm⁻² s⁻¹ (ref. 23), still two orders of magnitude less than the gas outflux from the volcanoes.

This inventory leads to an obvious but important conclusion. Only a small fraction of what is coming out of Io's volcanoes is escaping; the bulk of it is being recycled. There must be an atmospheric reservoir which maintains the gases until they have had a chance to recycle, perhaps by condensation to the surface on the nightside and subsequent chemical reactions.

Such an atmosphere must be protected from the co-rotating plasma in order to be stable. The comet-like interaction proposed by Cloutier *et al.*³ is still a promising approach, in which case the ionopause could be regarded as the atmospheric boundary beyond which all the neutrals are ionised and lost. Another attractive mechanism has been proposed recently by Kivelson *et al.*¹³ which invokes an intrinsic magnetic field of Io and a well developed magnetosphere which would provide a magnetic barrier to the co-rotating plasma and hence, shield the atmosphere.

Atmospheric constituents diffusing across the magnetopause or the ionopause boundary (~ 200 km in the upstream direction) will be subject to ionisation by hot electrons in the plasma torus and subsequent scavenging by the co-rotating plasma. According to solar-wind induced mass loss formalism of Michel¹⁴, mass loading of the co-rotating plasma will limit the maximum swept flux to 6×10^8 cm⁻² s⁻¹, about 1/40th of the magnetospheric wind itself. The Michel limit apparently does not apply since the

Table 1 Constituents detected in the vicinity of Io and in the associated plasma torus

Experiments	Measurement region	Constituent	Density (cm ⁻³)	Temperature (K)	Ref.
Ground-based telescopes	Cloud around Io	Na	10	$T_i \sim 10^4$	20
	Cloud around Io	S ⁺			21
Pioneer 10 UV experiment	Incomplete torus at Io's orbit	H			22
Pioneer 10 radio occultation experiment	Pre-sunset ionosphere	e	6×10^4		1
	Pre-sunrise ionosphere	e	9×10^3		2
Voyager UV experiment	Plasma torus remote sensing	S ²⁺	95	$T_i \sim 10^5$	10
		S ³⁺	55		
		O ²⁺	850		
Voyager plasma science experiment	Plasma torus $5.3 R_J < r < 6.3 R_J$ <i>in situ</i>	O ²⁺	Total ~ 2,000	$T_i \sim 10^5$	15
		S ²⁺ or O ⁺			
		S ⁺			
		S ₂ ⁺ or SO ₂ ⁺			
Voyager planetary radio astronomy experiment	Plasma torus <i>in situ</i>	e	2,000–4,000	$T_i \sim 10^5$	11
Voyager imaging experiment	Volcanoes	Dust umbrella			7, 8
Voyager IR experiment	A volcanic plume	SO ₂			19

needed flux input from Io into the plasma torus is comparable to the co-rotating plasma flux. Either Jupiter's magnetic field can impart sufficient momentum to sustain such a large flux input, or the co-rotation would cease at Io's orbit due to this mass loading. In any case the Io torus measurements from Voyager 1 imply that a gas flux of 10^{11} – 10^{12} cm⁻² s⁻¹ averaged over the globe could be lost from Io into the plasma torus.

Only the dayside atmosphere is subject to escape while the cold nightside and poles would serve as a sink of atmospheric gases. However, as shown in Fig. 1, even the dayside atmosphere will be completely protected over about half the Io's orbit when the dayside is towards the downstream direction. The scavenging of atmospheric gases is maximum at the eastern elongation and minimum at the western elongation.

The scavenging lifetimes of H, C, N, and O for the volatile escape can be estimated as follows. The atmospheric densities implied by the ionosphere are low enough that we can assume that each constituent is in diffusive equilibrium. In this case the depletion time, defined as the time it takes for the scavenging process to deplete the atmosphere by $1/e$ is given by $T_D = T_i \exp(z_b/H_i)$, where T_i is the lifetime against ionisation by plasma torus electrons, z_b is the magnetopause or ionopause altitude, and H_i is the scale height for the i th constituent. In the hot plasma torus region ($T = 10^5$ K) near Io, the measured electron density is $\sim 2,000$ cm⁻³ (refs 11, 15) which corresponds to an electron flux of 4×10^{11} cm⁻² s⁻¹. The energy of these electrons ~ 10 eV, is, however, marginal for ionisation of most constituents and this flux therefore represents an upper limit to the flux available for ionisation of atmospheric constituents. The cross-section for ionisation by low-energy electrons near the threshold is $\sim 10^{-16}$ cm² for most constituents. The resulting ionisation rate is 1 – 4×10^{-5} s⁻¹, probably closer to the lower value. Assuming an exospheric temperature of 1,000 K and $z_b = 200$ km, we obtain the atmospheric depletion times listed in Table 2. We caution that these lifetimes are based on the assumption that all neutrals diffusing across the boundary are ionised and lost. This is true if sufficient ionising flux is available and if the atmospheric column density above z_b does not exceed $\phi_{\max} T_i$, where $\phi_{\max}$ is the maximum flux that can be carried by the co-rotating plasma. If $\phi_{\max} = 10^{11}$ cm⁻² s⁻¹, the likely flux in view of the Voyager 1 data, the above assumption is valid for an SO₂ atmosphere with a surface density of $\leq 10^{11}$ cm⁻³ averaged over the globe. At higher densities the depletion times will be longer but the differential escape of atmospheric constituents will be maintained.

It is immediately apparent that the lifetimes are so short that the atmosphere must be replenished continuously. Note that the lighter gases escape more rapidly than the heavier ones, as a consequence of the diffusive separation assumed in our model. It is interesting, however, to compare these lifetimes to Io's orbital

period which is 1.77 days. Hydrogen, which has a depletion time shorter than the orbit period, will escape so rapidly that it has little chance of recycling at the surface. A large fraction of C, N, and O will also escape, but some will be recycled. Heavier elements, such as S, will be mostly recycled.

The implications of this differential loss on the evolution of Io's volatile content are significant. As a guide, it is instructive to consider the composition of gases ejected by the Earth's volcanoes. Rubey's¹⁶ inventory of volcanic gases from Kilauea and Mauna Loa includes following median values of volume percentages: H₂O(73.5), CO₂(11.8), N₂(4.7), SO₂(6.4), SO₃(2.3), S₂(0.2), H₂(0.4), CO(0.5), Cl₂(0.05), and A(0.2). Io may have had similar composition of gases coming out of its volcanoes in early geological history. Since Io most likely has a molten interior¹⁷, fresh material will be brought out to the surface by convective currents and hence, Io must have lost practically all of its hydrogen content over geological time through volcanic activity and subsequent escape. This could explain why no H₂O is observed on Io. Heavier gases CO₂ and SO₂ have much longer depletion times, but their escape would proceed through the rapid photodissociation of these constituents into atomic elements and their subsequent escape. The conspicuous absence of N(<10% of O) in the plasma torus¹⁵ and the absence of energetic nuclei of both C(<3% of O) and N(<7% of O) in the jovian magnetosphere near Io¹⁸ suggest that most of CO₂ and N₂ have also escaped from Io in addition to H₂O. Thus, NH₃ and N₂ previously suggested from the theoretical models^{3–6} must be ruled out. Remarkably, SO₂, the next most abundant gas in the Earth's volcanoes, has been observed above one of the hot spots on Io by the Voyager IR experiment¹⁹. This provides a very important milestone in the current state of evolution of Io and is understandable in terms of the above model. Most likely, SO₃, S₂, and argon have also been retained. Atomic oxygen and atomic sulphur should be present as photodissociation products. Some CO₂ may also be present as a minor constituent.

Table 2 Atmospheric depletion times on Io

Atmospheric constituent	Depletion time (d)
H	1.2
C	2.0
N	2.2
O	2.4
S	4.9
A	6.9
H ₂ O	2.6
NH ₃	2.5
CO ₂	8.3
SO ₂	20.0

Table 3 SO₂ atmosphere on Io

Surface Temperature <i>T</i> (K)	Vapour pressure ²⁵ <i>P</i> (bar)	Surface density <i>n</i> (cm ⁻³)
135	1.9×10^{-6}	1.4×10^{14}
130	5.8×10^{-7}	4.3×10^{13}
125	1.5×10^{-7}	1.2×10^{13}
120	3.5×10^{-8}	2.8×10^{12}
115	6.7×10^{-9}	5.6×10^{11}
110	1.0×10^{-9}	9.0×10^{10}
100	1.0×10^{-11}	1.0×10^9
90	2.1×10^{-14}	2.2×10^6

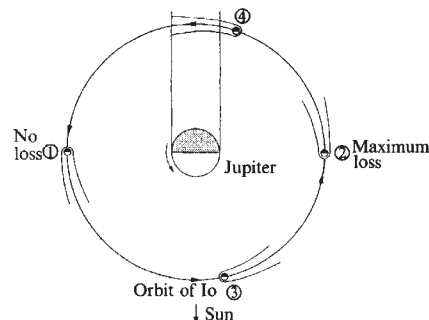


Fig. 1 The geometry of a magnetosphere protecting Io's atmosphere from Jupiter's co-rotating plasma at Io's orbit. Only the dayside would have a substantial atmosphere; hence the loss rate varies over Io's orbit. There is essentially no loss at the western elongation ①, while the scavenging rate above the magnetopause or the ionopause is maximum at the eastern elongation ②. Intermediate loss rates are expected at locations ③ and ④, which correspond to the location of Io in its orbit at the time of the Pioneer 10 radio occultation experiment and the Voyager 1 encounter, respectively.

In fact, the identification of gaseous SO₂ on Io, the evidence for solid SO₂ on the surface of Io²⁴, the measurement of sulphur and oxygen ions in the plasma torus^{10,15}, and the possible presence of SO₂⁺ in the plasma torus¹⁵ tempt us to consider SO₂ as the major gas in Io's atmosphere. Indeed, the upper limits to various candidate gases by the Voyager IR experiment¹⁹ are consistent with this scenario. We can further examine this possibility by testing whether an SO₂ atmosphere is consistent with the observed density and structure of Io's ionosphere and the relative S to O abundance in the plasma torus.

For this purpose, we consider the dusk or pre-sunset ionosphere measured by the Pioneer 10 radio occultation experiment¹, since in our model the dusk side ionosphere is protected while the dawn side ionosphere is facing the co-rotating plasma and is subject to compression or a complex intermixing of plasmas. The sunlit SO₂ atmosphere will be ionised by the solar UV radiation shortward of 1,000 Å producing SO₂⁺. In the photochemical equilibrium, the SO₂⁺ ions will be lost by recombination with electrons and the SO₂⁺ ion scale height is twice the SO₂ scale height. A fit to the electron density profile above the peak will require an exospheric temperature of 1,100 K. The ionosphere peak will occur at a level where the absorption optical depth $\tau = 1$. Requiring a $\tau = 1$ at 75 km and at solar zenith angle of 81° (the values corresponding to the Pioneer 10 radio occultation experiment), the SO₂ density is 2×10^9 cm⁻³ at this level. Assuming an SO₂ ionisation rate of 5×10^{-8} s⁻¹ (analogous to CO₂) and an SO₂⁺ recombination rate of $\sim 10^{-7}$ cm³ s⁻¹, the corresponding SO₂⁺ ion density at the peak is 2×10^4 cm⁻³, a factor of three less than the observed value. However, the SO₂ gas is a strong IR radiator, hence a cool SO₂ thermosphere ($T = 200$ K) analogous to the cool CO₂ thermospheres of Venus and Mars may not be unlikely. In this case, $\tau = 1$ will occur at a level where SO₂ density is 8×10^9 cm⁻³ and the corresponding peak SO₂⁺ density is 4×10^4 cm⁻³. Note that the exobase in this case is at ~ 105 km, not much above the ionosphere peak; thus the ion temperature could be substantially higher than the neutral temperature to match the observed ionosphere scaleheight. The peak electron density is still 30% less than the observed value which may be due to an uncertainty in the assumed ionisation cross-sections, recombination rates, and solar UV fluxes. Alternatively, the balance of ionisation may be provided by the precipitation of energetic electrons (~ 10 keV) or protons (~ 10 MeV) as suggested previously by McElroy and Yung⁴ for a molecular atmosphere. These uncertainties, however, should not introduce more than a factor of 10 error in the required SO₂ density at the ionosphere peak.

An extrapolation to the surface in a single slab thermosphere approximation yields an SO₂ surface density of 10^{11} – 10^{12} cm⁻³. The vapour pressure of SO₂ at 110–117 K appropriate near the evening terminator (J. Pearl, personal communication) will maintain this surface density (see Table 3). Although this is an order of magnitude estimate for the surface density, it is remarkable that an SO₂ atmosphere in thermal equilibrium with the surface is consistent with the observed ionosphere. In this case, the SO₂ surface density near the subsolar point ($T \sim 130$ K, J. Pearl, personal communication) could be as high as $\sim 10^{13}$ cm⁻³, which is consistent with a column of 0.2 cm atm (surface density $\sim 5 \times 10^{12}$ cm⁻³) at 1 p.m. local time on Io observed by the Voyager IR experiment¹⁹.

The relationship between the torus ion densities and the atmospheric S and O abundance will require a photochemical model, but a rough estimate can be made as follows. Total sulphur ion densities add up to ~ 150 cm⁻³, whereas total oxygen ion density is 850 cm⁻³ (ref. 10). Assuming that all O and S atoms above the magnetopause boundary are converted to ions, this implies a density ratio [O]/[S] = 3 at 200 km. For a 1,000 K exosphere, this scales to comparable O and S densities below the ionosphere peak at 75 km which is what one would expect since most of the SO₂ dissociation would take place within one scale height below the ionosphere peak.

In conclusion, a stable SO₂ atmosphere on Io shows promise if the volcanic activity is continuous. If, however, the volcanoes stop erupting, the atmosphere will be depleted at a rapid rate. Note that the atmosphere must be protected in order to be stable, either by a magnetosphere or by slowing down of the co-rotating plasma in a comet-like interaction. Both of these cases lead to a differential escape of gases which is required by the present state of Io's atmosphere.

I thank J. Pearl, R. Hanel, T. Johnson, E. Stone, G. Siscoe, and D. Hunten for discussions and previews of Voyager 1 results. This work was supported by NASA Contract no. NAS7-100.

Received 27 June; accepted 24 July 1979.

- Kliore, A. J., Cain, D. L., Fjeldbo, G., Seidel, B. L. & Rasool, S. I. *Science* **183**, 323 (1974).
- Kliore, A. J. *et al. Icarus* **24**, 407 (1975).
- Cloutier, P. A., Daniell, R. E., Dessler, A. J. & Hill, T. W. *Astrophys. Space Sci.* **55**, 93 (1978).
- McElroy, M. B. & Yung, Y. L. *Astrophys. J.* **196**, 227 (1975).
- Sinton, W. M. *Icarus* **20**, 284 (1973).
- Gross, S. H. & Ramanathan, G. V. *Icarus* **29**, 493 (1976).
- Smith, B. A. *et al. Science* **204**, 951 (1979).
- Morabito, L. A., Synnott, S. P., Kupferman, P. N. & Collins, S. A. *Science* **204**, 972 (1979).
- Johnson, T. V., Sagan, C. & Soderblom, L. A. *Nature* **280**, 746–750 (1978).
- Broadfoot, A. L. *et al. Science* **204**, 979 (1979).
- Warwick, J. W. *et al. Science* **204**, 995 (1979).

- Eviatar, A., Siscoe, G. L. & Mekler, Y. *Icarus* (in the press).
- Kivelson, M. G., Slavin, J. A., Southwood, D. J. *Science* (in the press).
- Michel, F. C. *Planet. Space Sci.* **19**, 1580 (1971).
- Bridge, H. S. *et al. Science* **204**, 987 (1979).
- Rubey, W. W. *Bull. geol. Soc. Am.* **62**, 1111 (1951).
- Peale, S. J., Cassen, P. & Reynolds, R. T. *Science* **203**, 892 (1979).
- R. E. Vogt *et al. Science* **204**, 1003 (1979).
- Pearl, J. *AGU Spring Meet.* Washington D.C., (1979).
- Bergstralh, J. T., Young, J. W., Matson, D. L. & Johnson, T. *Astrophys. J.* **211**, L51 (1977).
- Kupo, I., Mekler, Y. & Eviatar, A. *Astrophys. J.* **205**, L51 (1976).
- Judge, D. L. & Carlson, R. W. *Science* **183**, 317 (1974).
- McNutt, R. L. Jr, Belcher, J. W., Sullivan, J. D., Bagenal, F. & Bridge, H. S. *Nature* **280**, 803 (1979).
- Fanale, F., Brown, R. H., Cruikshank, D. T. & Clark, R. N. *Nature* **280**, 761–763 (1979).
- Dean, J. A. (ed.) *Lange's Handbook of Chemistry* (McGraw-Hill, New York, 1973).

Significance of absorption features in Io's IR reflectance spectrum

Fraser P. Fanale*, R. Hamilton Brown†, Dale P. Cruikshank† & Roger N. Clarke‡

* Space Science Division, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91103

† Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822

‡ Massachusetts Institute of Technology, Department of Earth and Planetary Sciences, Cambridge, Massachusetts 02139

Absorption features in Io's reflectance spectrum suggest that frozen SO₂ molecules are present on the surface of Io as free frost; this may have important implications for Io's atmosphere.

Io's deep 4.1- μm absorption may be due to frozen SO₂ molecules incorporated in or adsorbed on sulphur allotropes and other surface material, and present as free frost. Shallower absorption features near 3.8 μm and 2.6 μm may be due to H₂S trapped in surface material. We show here that the presence of SO₂ may have important implications for Io's atmosphere and transient phenomena, and suggest that any abundant metal-containing compounds are spectrally bland in the 2.0–5.0 μm range, such as halides or sulphides.

Among Io's many remarkable properties is a reflectance spectrum showing a steep dropoff in the UV, a high visible and near-IR albedo, and the absence of major absorptions in the 1.0–3.0 μm region^{1,2}. It has been suggested^{3–5} that those properties could be explained if the surface were largely covered with an anhydrous mix of sulphur and a variety of salts, mainly sulphates. Fanale *et al.*^{3,4} further argued that such a unique surface could have formed by endogenic activity, specifically the percolation of aqueous solutions to Io's surface from a warm or hot interior, followed by H₂O loss to space^{3,4}. Wamsteker *et al.*⁶ also argued that elemental S was a major constituent and pointed out that the non-sulphur component had a visible spectrum resembling that of Saturn's rings.

The so-called 'evaporite' hypothesis^{3,4} in which aqueous solutions are the carrier requires an efficient process of subsequent surface dehydration because IR spectra of Io^{7–9} place increasingly severe limits^{2,9} on the abundance of H₂O in any form in the optical surface. Charged particle sputtering has been suggested¹⁰ as a surface dehydration mechanism. The presence of a major absorption near 4.1 μm has been reported and its interpretation is considered here.

Voyager observed intensive current volcanic activity at many sites on Io^{11,12}, and just before Voyager's encounter, Peale *et al.*¹³ had predicted extensive tidal heating of Io by Jupiter. Fanale *et al.*^{3,4} had considered only radioactive decay as a current heat source on Io which led to the prediction that Io had undergone extensive devolatilisation but would currently possess a 400 km solid shell overlying molten silicate⁴. Although they speculated that ordinary magmatism might occur at the surface, the migration of aqueous solutions seemed more compatible with their thermal model—hence the postulated carrier role for H₂O. Now, however, with theory¹³ and observation^{11,12} suggesting a very shallow solidus and rapid surface renewal¹⁴ it seems unnecessary to postulate any current efficient dehydration mechanism. Io-forming material may have originally contained substantial amounts of bound H₂O^{3,4} and a greater role for H₂O earlier in Io's history is still an exciting speculation, though beyond the scope of this note.

With the dehydration problem mitigated, it remains to explain the deep 4.1- μm band without violating other constraints imposed by Io's spectrum or other observations.

Figure 1 shows Io's 0.3–5.2 μm reflectance spectrum and those of some possible Io surface constituents. Our laboratory

spectra from 2.4 to 5.3 μm were obtained using a spectrometer constructed from two main components: (1) a chopped quartz halogen light source run in a low voltage d.c. mode, and (2) an indium antimonide photometer having a two-segment circular variable filter—the same instrument used to record the satellite spectrum^{7,8}. The spectra were taken using sublimed sulphur as a reflectance standard. The circular variable filter consisted of two segments, the first of which had a wavelength range of 2.37–4.52 μm with $\Delta\lambda/\lambda \sim 0.01$. The second had a wavelength range of 3.9–7.7 μm with $\Delta\lambda/\lambda \approx 0.025$. The laboratory spectra were recorded on a chart recorder, then digitised and processed by a computer. Processing amounted to matching the two halves of the spectrum and then ratioing each spectrum to the composite sulphur reflectance standard. Wavelength calibration was obtained from distinct absorptions in the known bands in the KNO₃ reflectance spectrum. In cases where samples were dehydrated, evacuated containers with sapphire windows were used, otherwise the samples were run in Petri dishes open to the air. The laboratory spectra in the 0.33–2.5 μm region were taken with a spectrogoniometer and environment chamber using a Halon standard and similar computer processing (Halon is a white reflectance standard for the 0.3–2.5 μm spectral region that does not absorb water). Again, the same instrument used in obtaining the Io spectrum^{2,15} was used in the laboratory.

Spectroscopic arguments against major coverage of Io's surface by many candidate surface materials including rock-forming silicates, many frosts, certain salts, and other phases have been reported previously^{3–5}. When the deep 4.1- μm band was first observed it was speculated that it might be an overtone band of an anion of a salt such as a carbonate or a nitrate^{7,9} since a major absorption at 4.0–4.1 μm is present in the spectra of both groups of minerals. However, carbonates and nitrates also exhibit other bands—some almost as deep as the band around 4 μm —in the region 3.5–3.6 μm (ref. 16) which are not present in Io's spectrum (see Fig. 1). Also, Kieffer *et al.*¹⁷ have recently shown that altering the grain size and temperature of the carbonates and nitrates does not significantly change the ratio of the band depth at 3.6 μm to that at 4.1 μm . With these possibilities largely voided, we reconsidered the possibility that elemental sulphur alone might explain Io's deep 4.1- μm absorption. Ordinary 'flowers of sulphur' is S₈ and there are no major absorptions for S₈ in the 2.0–5.0 μm region⁷. In fact, we use S₈ as a near IR reflectance standard. However, Nelson and Hapke¹⁸ have shown that S₂, S₃ and S₄ have very different visible and UV spectra, and that a mixture of these polymers can explain some of Io's reflectance characteristics in the visible and UV better than pure S₈, a mix including polymeric chains such as S₄ and S₂. Furthermore, they stated that heating S₈ to >160 °C and quenching produces a mix containing S₂, S₃, S₄ and possibly S₆. Thus to check for possible 4 μm absorption, we prepared an altered sulphur sample by heating sulphur in air until it reached a temperature of ~200 °C. It was then quenched in liquid nitrogen, ground in liquid nitrogen and kept under liquid nitrogen until the spectrum was run. At run time the sample was transferred to a sapphire window container which was subsequently evacuated. Several spectra were taken at ~5 min intervals over a period of 75 min as the sample warmed to room temperature. The processed spectra show a large water band at

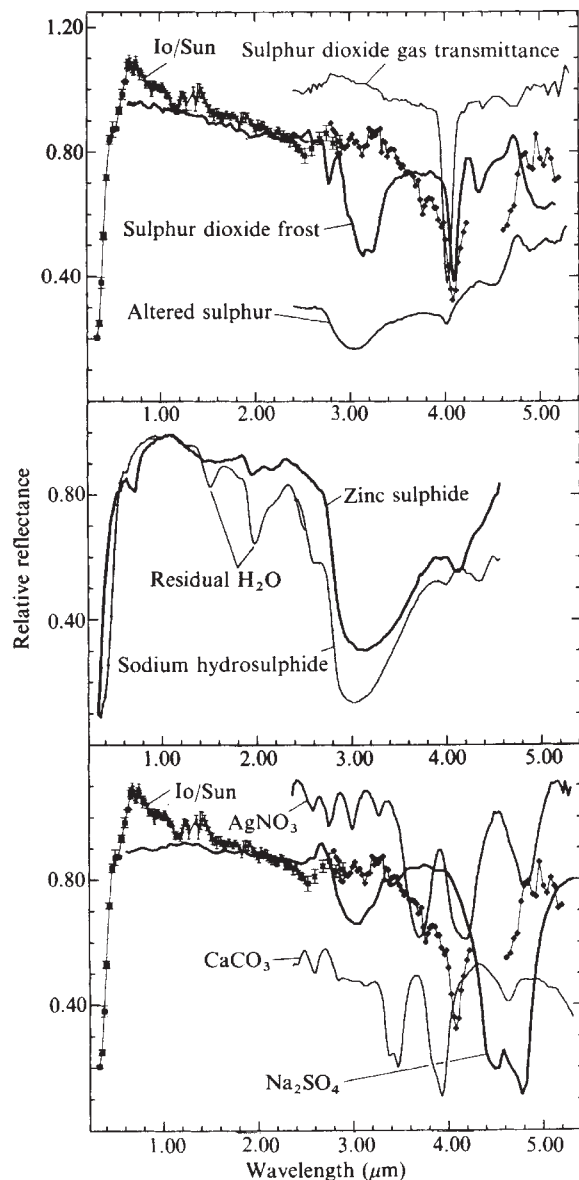


Fig. 1 The reflectance spectrum of Io (leading side, except the Cruikshank 1979 data⁸ from 2.9 to 5.2 μm which are mixed orbital phases) is compared with various laboratory spectra. The Io composite spectrum is from Clark and McCord² and contains data from McFadden *et al.*¹⁵ (0.325–0.65 μm), Clark and McCord² (0.65–2.5 μm), Pollack *et al.*⁹ (2.5–3.0 μm), and Cruikshank⁸. The Na_2SO_4 spectrum (H. H. Kieffer, personal communication) is plotted as reflectance. All other laboratory spectra are from this study. The SO_2 frost, ZnS, NaSH and Na_2SO_4 are plotted as reflectance, whereas the other laboratory data are scaled for better clarity. The Io data are scaled to 1.0 at 1.02 μm (the geometric albedo of Io at 1.02 μm is approximately 0.90). The absorption bands at $\sim 3 \mu\text{m}$ in the laboratory data are due to a small amount of water in the samples.

3.0 μm due to contamination during handling of the cold sample, and three bands at 4.03, 4.4 and possibly at 5.0 μm , with the 4.03- μm band being the sharpest and most prominent. The shape of the spectrum, as well as the apparent band depths, changed radically while the sample was warming. The final (room-temperature) spectrum is shown in Fig. 1. This suggested contamination by a spectrally active species such as H_2S or SO_2 as an alternative to the intended generation of polymeric sulphur. Further, it was noted that the spectrum of reagent grade ZnS (Fig. 1) exhibited a band in essentially the same position and it would seem difficult to attribute any of the ZnS bands to causes other than contamination, because the Zn–S fundamental occurs¹⁹ between 306 and 280 cm^{-1} . Also a natural sphalerite sample that was run failed to show the same band. It was also noted that SO_2 gas has a deep absorption near 4 μm (ref. 20). Finally SO_2 gas has been observed by the Voyager IRIS experiment^{21,22}.

All these considerations at first led us to suspect that gaseous atmospheric constituents might be contributors to Io's 3.0–5.2 μm spectrum. However, data of Pierson *et al.*²⁰ show that the production of a 30% absorption near 4 μm requires 1.6 cm amagat of SO_2 , which is approximately 10^4 times the limit based on Pioneer 10 occultation results^{23,24} and what is known of the ionising plasma flux near Io²⁵. Hydrogen sulphide has bands at 2.6 and 3.8 μm , near possible weak features in Io's spectrum, but, again, even a 5% atmospheric absorption at either of these wavelengths requires >2 cm amagat²⁰. Note that, based on Io's 0.8–2.7 μm spectrum, Fink *et al.*²⁵ set limits for NH_3 , CH_4 , N_2O and H_2S in Io's atmosphere of 0.12, 0.12, 0.4 and 24 cm amagat, respectively. Furthermore, the gaseous SO_2 absorption band occurs at 4.03 μm , and after careful reexamination of our calibration procedures, we concluded that the difference between 4.03 and 4.08 μm , where the salient major band centre is observed on Io, is significant.

The spectrum of H_2S frost has already been reported by Kieffer *et al.*¹⁷, and it exhibits a deep absorption somewhat shortward of 4.0 μm and several other major absorptions. We therefore obtained an IR reflectance spectrum of SO_2 frost. The 2.4–5 μm portion of the SO_2 frost spectrum was obtained from a sample grown on a flat metal surface attached to a LN_2 -filled cold finger in an evacuated chamber with a sapphire window. The gas was slowly admitted from a lecture bottle via a line that was cold-trapped to remove any residual water vapour. During observations of the spectrum, the chamber was kept in a vacuum with the mechanical pump isolated by a cold trap. The frost was maintained at <100 K. The sulphur reflectance standard was obtained in the same conditions. The 0.65–2.5 μm portion of the spectrum was obtained by manufacturing SO_2 gas by reacting concentrated sulphuric acid and sodium sulphite and growing the frost on a cold (<160 K) metal surface. The SO_2 gas was dried by bubbling the gas through concentrated sulphuric acid, calcium carbonate, and a cold trap. In both methods of frost growth, the resultant samples were very fine grained ($\leq 30 \mu\text{m}$).

The SO_2 frost spectrum (Fig. 1) was found to exhibit a major absorption at 4.08 μm , corresponding to the SO_2 gas band at 4.03 μm with the expected shift to longer wavelength in the solid state. Other bands observed are minor, attributed to minor H_2O contamination, or compatible with the 0.7–5.0 μm spectrum of Io. Although Io exhibits a comparably pronounced absorption to that provided by our frost sample, the fact that we observed a deep broad absorption in the ground, altered sulphur suggests that SO_2 might strongly affect the major absorption as occluded or dissolved SO_2 as well as free frost. It is crucial, nonetheless, to assess the compatibility of free frost *per se* with Io's atmospheric properties. It is difficult to assess whether SO_2 could survive in the rough competition between surface renewal and destruction which must occur on Io¹⁴, but easy to show that SO_2 frost might be physically stable in equilibrium. At 100 K, a reasonable cold-trapping temperature for Io, the equilibrium vapour pressure of SO_2 is 3×10^{-9} torr. This is equivalent to a surface number density of $4 \times 10^8 \text{ cm}^{-3}$, in accord with current atmospheric models²⁴. Thus, given higher SO_2 concentrations at certain sites, it would seem reasonable to expect frost formation. Conversely, if SO_2 frost were widespread, it might, together with the occultation limit, imply that SO_2 is the main atmospheric constituent. If so, this would in turn imply higher atmospheric temperatures than transient H_2S or other constituents²⁵. The H_2S vapour pressure is much higher than that of SO_2 and violates the occultation limit by three orders of magnitude at 95 K. Remembering also that H_2S gas must be present at the level of >2 cm amagat to influence strongly the spectrum, we conclude that only H_2S trapped in surface material could strongly affect the 3.0–5.0 μm spectrum. Such occluded H_2S could, however, help to explain the ancillary features at 3.8 and 2.6 μm since its gaseous spectrum shows absorptions there, and not elsewhere in the region of interest, as reported by Pierson *et al.*²⁰.

With most of the identifiable features accounted for, it is not necessary for any compounds on Io's surface which may contain metallic cations to be spectrally active. In contrast, any abundant

phases must be spectrally bland in the region 1.0–5.0 μm (except possibly in the noisy ambiguous region 4.2–5.2 μm). With extensive volcanic activity, release of volatile metals may be expected. Regardless of whether Na and other constituents are sputtered from Io's surface or added directly to space from volcanoes, some metallic cations are liable to find their way to the surface. At low oxygen partial pressures, terrestrial volcanic sublimates are often dominated by alkali metals in the form of halides^{26–28}. It is common knowledge that halides are spectrally bland in the near IR and would satisfy our criteria, now that the major features are explained, but Io may have 'run out' of halogens by this point in its active devolatilisation history. Other common sublimate phases include sulphates, often but not always, as the result of oxidation by air^{27–29}. On Io sulphides may be more likely. As Fig. 1 shows, sulphates have a huge band at 4.2–4.5 μm and it is questionable how much 'room' there is for such a band in that region of Io's spectrum. Sulphides and hydrosulphides are spectrally bland in the 3.0–5.0 μm region and very compatible with Io's spectrum otherwise (Fig. 1). Therefore they may be more likely candidates to harbour the cations.

The great depth of the 4.1- μm band suggests that SO_2 in some form in solid surface material must be almost ubiquitous. The concentration of the brightest material near the equator puzzles us, but may be related to a greater ion bombardment flux near the poles or proximity to gas sources. If SO_2 could condense anywhere on Io's surface as a free frost, then it may prove worthwhile to reconsider Binder and Cruikshank's³⁰ and

Sinton's³¹ suggestions that post-eclipse brightening could be caused by a frost. A straightforward experimental test would be to look for a sharp correlation of both rotational and post-eclipse darkening at 4.08 μm with visual brightening. Other peculiar properties of Io such as the discrepancy between its 10- μm and 20- μm temperatures may also be related to frost phenomena. Thus, many of Io's peculiar properties may be a result not only of the occurrence of widespread current volcanism, but also of the specific physical properties of the volatiles that are issuing from the volcanoes. Finally, as the resemblances between the visible and near UV spectra of the 'non-sulphur' spectrum of Io and Saturn's rings and Europa has been considered, the possibility that the presence of non- H_2O frosts or dissolved gases may affect the properties of the latter two or Ganymede or Callisto should be investigated.

This paper represents one phase of research carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, under contract NAS 7-100 and the auspices of the Planetology Program. This research was primarily conducted at the Institute for Astronomy of the University of Hawaii where F. P. F. was a visiting scientist. R. N. C. was supported by NASA grants NSG 7323 and NSG 7312 and D. P. C. and R. H. B. by NASA grant NGL 12-001-057. F. P. F. thanks the Institute for Astronomy personnel, especially T. B. McCord, for scientific support. Technical discussions with M. Gaffey, G. Hunt, T. V. Johnson, H. H. Kieffer, L. Lebofsky, T. B. McCord, D. B. Nash, R. M. Nelson, J. B. Pollack, P. Russ III, L. A. Soderblom, and the Voyager IRIS team are acknowledged.

Received 23 May; accepted 20 July 1979.

- Johnson, T. V. & Pilcher, C. B. *Planetary Satellites* (ed. Burns, J.) 232 (University of Arizona Press, 1977).
- Clark, R. N. & McCord, T. B. *Icarus* (in the press).
- Fanale, F. P., Johnson, T. V. & Matson, D. L. *Science* **186**, 922 (1974).
- Fanale, F. P., Johnson, T. V. & Matson, D. L. *Planetary Satellites* (ed. Burns, J.) 379 (University of Arizona Press, 1977).
- Nash, D. B. & Fanale, F. P. *Icarus* **31**, 40–80 (1977).
- Wamsteker, W., Kroes, R. L. & Fountain, J. A. *Icarus* **23**, 417 (1974).
- Cruikshank, D. P., Jones, T. J. & Pilcher, C. B. *Astrophys. J. Lett.* **2**, L89 (1978).
- Cruikshank, D. P. *Icarus* (submitted).
- Pollack, J. B. *et al. Icarus* **36**, 271 (1978).
- Pilcher, C. B. *Icarus* **37**, 559–574 (1979).
- Morabeto, L. A., Synnott, S. P., Kupferman, P. N. & Collins, S. A. *Science* **204**, 972 (1979).
- Smith, B. A. *et al. Science* **204**, 951 (1979).
- Peale, S. J., Cassen, P. & Reynolds, R. T. *Science* **203**, 892 (1979).
- Johnson, T. V., Soderblom, L. A. & Sagan, C. *Nature*, **280**, 746–750 (1979).
- McFadden, L. A., McCord, T. B. & Bell, J. *Icarus* (in the press).
- Hovis, W. A. *Appl. Opt.* **5**, 245–249 (1964).

- Kieffer, H. H., Pleskot, L. K. & Smythe, W. *Bull. Am. astr. Soc.* **10**, 580 (1978).
- Nelson, R. M. & Hapke, B. W. *Icarus* **36**, 404 (1978).
- Liese, H. C. in *Infrared and Raman Spectroscopy of Lunar and Terrestrial Materials* (ed. Karr, C. Jr) 197 (Academic, New York, 1975).
- Pierson, R. H., Fletcher, A. N. & St. Clair Gantz, E. *Analyt. Chem.* **28**, 1218 (1956).
- Pearl, J. C. *et al. Nature* **280**, 755–758 (1979).
- Kumar, S. *et al. Nature* **280**, 758–760 (1979).
- Kliore, A. O., Cain, D. L., Fjelbo, G. & Seidel, B. L. *Science* **183**, 232 (1974).
- Kliore, A. J., Fjelbo, G., Seidel, B. L., Sweetnam, D. N., Sesplankis, T. T. & Woiceshyn, P. M. *Icarus* **24**, 407 (1975).
- Johnson, T. V., Matson, D. L. & Carlson, R. W. *Geophys. Res. Lett.* **3**, 293 (1976).
- Fink, U., Larson, H. P. & Gautier, T. N. *Icarus* **27**, 439 (1976).
- Naughton, J. J., Lewis, V. A., Hammond, D. & Nishimoto, D. *Geochim. cosmochim. Acta* **38**, 679 (1974).
- Naughton, J. J., Greenberg, V. A. & Gouguel, R. J. *Volcan. Geotherm. Res.* **1**, 149 (1976).
- Naughton, J. J., Lewis, V., Thomas, D. & Finlayson, J. B. *J. geophys. Res.* **80**, 2963 (1975).
- Binder, A. B. & Cruikshank, D. P. *Icarus* **3**, 299 (1964).
- Sinton, W. M. *Icarus* **20**, 284 (1973).

Spectral evidence for sublimates and adsorbates on Io

Douglas B. Nash & Robert M. Nelson

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91103

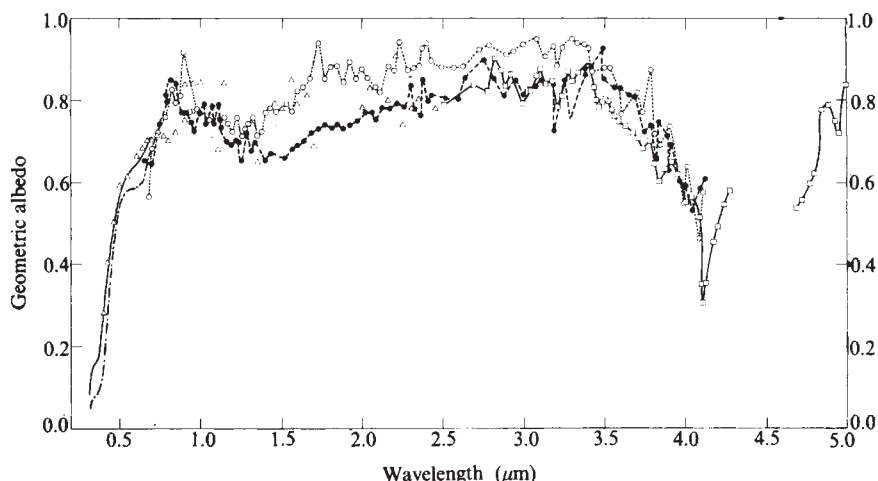
The reflectance spectrum of Io can be explained by a surface layer consisting of fine-grained particles of sublimated alkali sulphides and sulphur on which H_2S and SO_2 are adsorbed.

THE discovery by Voyager 1 of Io volcanoes, and thus active degassing of Io's interior, suggests that deposition from a vapour phase may be the dominant process producing the composition and texture of Io's optical surface. We report here our first results from laboratory studies regarding sublimates and adsorbates, materials not seriously considered in previous attempts to explain Io's spectral reflectance^{1–4}. We believe that sublimation and adsorption considerations open a new and fruitful area of

study for the interpretation of Io's optical properties. We conclude that Io's spectrum from 0.25 to 5.0 μm can be explained by an optical surface layer consisting of fine-grained particles of sublimated alkali sulphides and free sulphur on which is adsorbed a substantial component of H_2S and/or SO_2 .

Io's spectral geometric albedo (Fig. 1) is unusual in that it is strongly absorbing in the UV and very reflecting in the IR⁵. The important features in Io's spectral geometric albedo are (from UV to IR) a downturn shortward of 0.34 μm which when combined with spacecraft UV data suggests a band at 0.29 μm , a sharp upturn longward of 0.38 μm on the leading side and 0.40 μm on the trailing side, an absorption feature at 0.56 μm , a possible broad band at 1.35 μm , weak bands at 2.55, 3.45, 3.8, 3.9 μm and a strong band at 4.1 μm ^{4,6–13}.

Fig. 1 Spectral geometric albedo of Jupiter's satellite, Io, 0.26–5.0 μm based on previous studies $\theta = 135^\circ$ (solid line) ref. 4; $\theta = 307^\circ$ (chain line) ref. 4; Δ , refs 6, 7, data below 0.35 μm not included due to filter problems; $\circ$, ref. 8, leading side; $\bullet$, ref. 8, trailing side; $\square$, ref. 9 mixed. The leading side albedos ($0^\circ < \theta < 180^\circ$) have been normalised to 0.622 at $\lambda = 0.56 \mu\text{m}$, the trailing side albedos ($180^\circ < \theta < 360^\circ$) have been normalised to 0.58 at $0.56 \mu\text{m}$ following the discussion in refs 10–12.



To interpret these spectral features we measured the hemispherical and bidirectional spectral reflectance of typical sublimate phases in varied temperature and particle (proton beam) irradiation conditions. We used phases containing elements that are found in Io's extended exosphere and thus which may originate from Io's surface¹⁴. The materials studied were Na_2S , NaHS , K_2S , and mixtures thereof with free sulphur. We used reagent grade materials and, because of their extreme hygroscopic nature, prepared them in dry-nitrogen filled glove bags fitted to all the instruments used. The only exception to this procedure was in the 2.5–5.0 μm spectrophotometer runs where the samples were briefly (~ 10 s) exposed to low-humidity room air during transfer. This exposure was not considered significant because of subsequent sample drying by purge air in the instrument.

The hemispherical spectral reflectance relative to MgO , in the range 0.24–2.5 μm , was measured for the powdered samples ($\leq 100 \mu\text{m}$ particle size) with a Beckman DK-2A spectrophotometer. The bidirectional reflectance 2.5–5.0 μm for all three samples was measured relative to an aluminium mirror using a Perkin-Elmer model 180 spectrophotometer modified to take horizontal upright samples and purged with air filtered of substantially all H_2O and CO_2 .

Because of the very strong water bands in this spectral region, we repeated the 2.5–5.0 μm spectra after the powders were heated in a vacuum oven at 170°C for 24 h.

The results of the 2.5–5.0 μm spectrophotometry are shown in Fig. 2 for the freshly powdered samples and Fig. 3 for the vacuum oven dried powdered samples. The upper portions of Figs 2 and 3 show the ordinate displaced spectral geometric albedos of Io. Note the apparent disagreement in wavelength of the band centres of the reported Io spectra^{8,9}.

We find that Na_2S , K_2S , and NaHS all have features in the 3.9–4.1 μm region at the location of the strong Io bands. The band centres for the laboratory samples are at 3.96 and 4.05 μm for Na_2S , 3.93 and 4.11 μm for K_2S , and 3.93 and 4.05 μm for NaHS . Furthermore, both K_2S and Na_2S have a feature at $\sim 2.55 \mu\text{m}$, and NaHS has features at 3.4 and 3.8 μm . Io has band minima at all of these wavelengths.

The drying procedure did little to diminish the depth of the adsorbed water band at 3.0 μm ; however, it did alter dramatically the depth of the non-water related bands at $\sim 4.0 \mu\text{m}$. We suspect that this may represent surface phase alterations in addition to desorption of water.

A possible explanation of the observed $\sim 4.0 \mu\text{m}$ band structure in these sulphides is the presence of H_2S or SO_2 as an adsorbate phase. As clearly shown in Fig. 3, all three sulphides yield IR bands near 4.0 μm . The S–H stretching bond frequencies are well known to produce bands near this wavelength¹⁵. The $\nu_1 + \nu_3$ band of SO_2 gas is also at 4.0 μm (ref. 15). The precise positions of these bands depends on the details of the molecular structure¹⁶. We found that an odoriferous gas was evolved

when these materials were dried, either by vacuum oven heat or dry purge air in the spectrophotometer. As the H_2O adsorbate was driven off, the 4.0- μm bands became stronger. We interpret this effect as being due to formation of H_2S or SO_2 by chemical reaction with the matrix alkali sulphide and the retention of the gas on the surface of the sulphide particles as a chemisorbed phase.

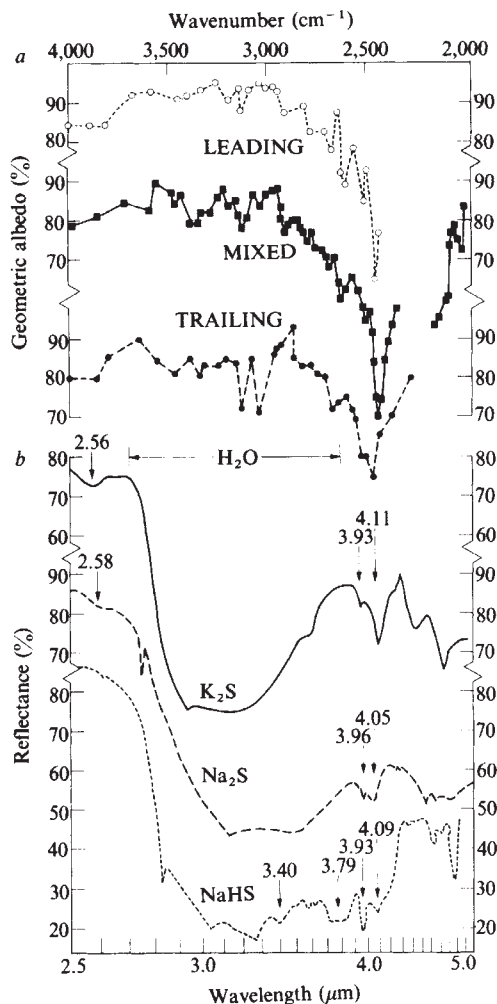


Fig. 2 a, The ordinate-displaced spectral geometric albedos of Io from ref. 8 ($\circ$, $\bullet$) and ref. 9 ($\blacksquare$). Note the apparent disagreement in the positions of the band centres between the two studies. b, The bidirectional reflectances of the selected sublimates at a phase angle of $\sim 20^\circ$ at room temperature.

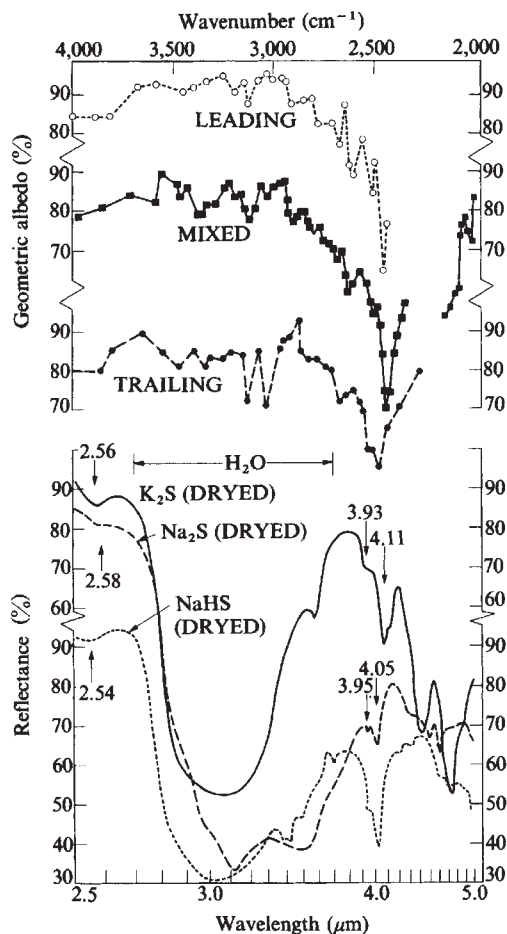


Fig. 3 *a*, The ordinate displaced spectral geometric albedos of Io from ref. 8 (○, ●) and ref. 9 (■). *b*, The bidirectional reflectances of the selected sublimates at room temperature after 24 h in a vacuum oven at 170 °C.

Because of the strong similarity of this band structure to Io's 4.0-μm band, we suggest that adsorbed gas may be its cause on Io. H₂S and SO₂ are strongly polarised molecules as is H₂O which produces strong IR bands in terrestrial powders even though present only as an adsorbed component (see ref. 17). H₂S is unstable as a frost on Io¹⁸ but can reasonably be expected

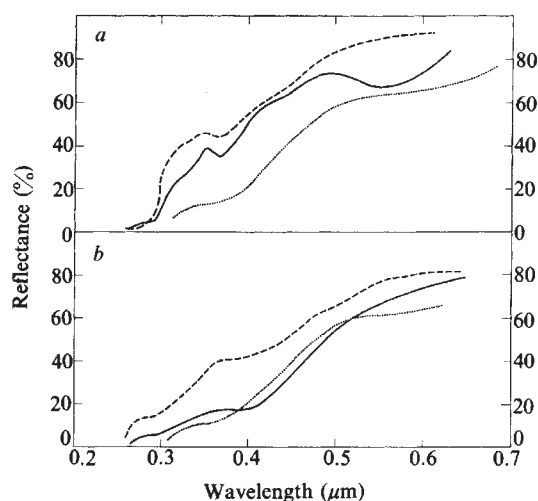


Fig. 4 *a*, Spectrum of Na₂S before (dashed line) and during (solid line) low-energy proton irradiation at $T = 170$ K. *b*, Spectrum of NaHS before (dashed curve) and during (solid line) low energy proton irradiation $T = 300$ K. The Io spectrum⁴ is shown for comparison (dotted line) in both *a* and *b*.

as an adsorbate phase on the fine-grained, and thus high-specific-surface, sublimate particles on Io's surface. We are conducting more detailed experiments to test this hypothesis.

To examine possible effects of jovian charged particle flux at Io on alkali sulphides, the powdered samples of Na₂S and NaHS were introduced into a vacuum system at room temperature (see ref. 19) where they were subjected to a dose (2×10^{15} H⁺ cm⁻²) of 5 keV protons. Bidirectional reflectance relative to MgO was measured (0.28–0.70 μm) before and during irradiation. This irradiation treatment was repeated for Na₂S at 170 K.

Before irradiation the NaHS had a shoulder at 0.4–0.6 μm and a downturn at 0.35 μm. The Na₂S had bands at 0.30 μm and an increase in reflectance between 0.36 to 0.55 μm. The spectra of Na₂S and NaHS before and during irradiation are shown in Fig. 4. Both samples were darkened by the proton irradiation, causing the NaHS to more closely match Io's spectrum and albedo in this wavelength range. The irradiation also caused a band in Na₂S to form at 0.56 μm, where Io has a distinct feature. The proton irradiation did not cause any spectral changes in the IR spectra of Na₂S or NaHS.

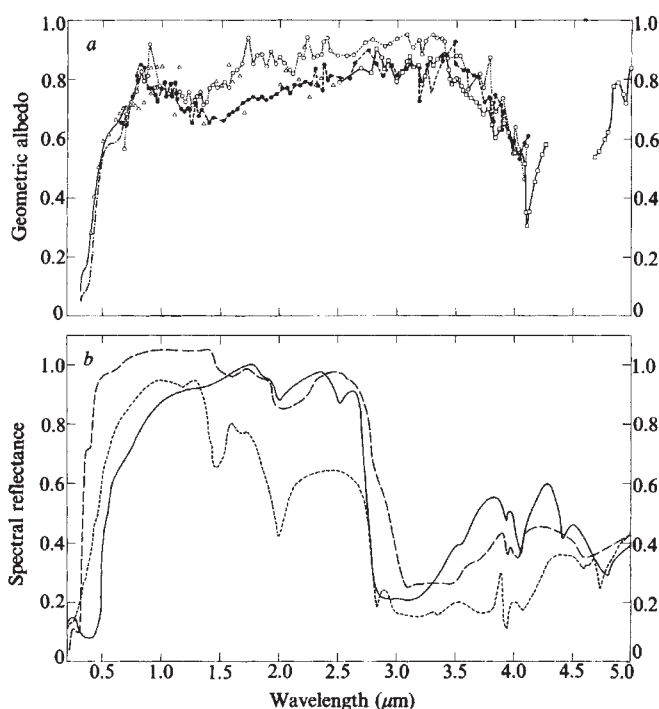


Fig. 5 *a*, Spectral geometric albedo of Io 0.28–5.0 μm. $\theta = 135^\circ$ (solid line) ref. 4; $\theta = 307^\circ$ (chain line) ref. 4; Δ , refs 6, 7 data below 0.35 μm not included due to filter problems; ○, ref. 8, leading side; ●, ref. 8, trailing side; □, ref. 9, mixed. *b*, Spectral reflectance of the selected sublimates 0.24–5.0 μm. The deep bands at 1.5, 2.0 and 3.3 μm are due to water in the laboratory samples. Na₂S (dashed line); K₂S (solid line); NaHS (dotted line).

Figure 5 shows the spectra of the three sublimates for the complete wavelength range 0.25–5.0 μm. The Io spectral geometric albedo from Fig. 1 is shown for comparison. In addition to matching the principal absorption features in Io's spectrum quite well, all three sublimates (when water free) have the general continuum shape of Io's spectrum.

Conclusions

We conclude that sublimates and adsorbates are important, if not the dominant, phases causing Io's observed spectral reflectance in the range 0.25–5.0 μm. In particular, the compounds of Na₂S, K₂S and NaHS have absorption bands at all the key wavelengths in the UV and visible range. Mixtures of sulphides and free sulphur are being studied in the laboratory in efforts to produce a refined compositional analysis, especially to fit the overall continuum of Io's spectrum. Furthermore, we believe that adsorbed gases if not frosts are important phases affecting the IR spectrum of Io's surface. The adsorptivity of

various sulphur and other compounds and the physical and optical effects of adsorbed gases such as H_2S and SO_2 thereon need to be studied as possible strong contributors to Io's spectral reflectance and other surface properties.

We thank Dennis L. Matson for helpful conversations, and

George Rossman and Lois Taylor for technical assistance. This work represents the results of one phase of research carried out at the California Institute of Technology JPL under contract NAS 7-100, sponsored by NASA. R.M.N. is an NRC resident research associate.

Received 1 June; accepted 12 July 1979.

1. Fanale, F. P., Johnson, T. V. & Matson, D. L. *Science* **186**, 922-925 (1974).
2. Nash, D. B. & Fanale, F. P. *Icarus* **31**, 40-80 (1977).
3. Wamsteker, W., Kroes, R. L. & Fountain, J. A. *Icarus* **23**, 417-424 (1974).
4. Nelson, R. M. & Hapke, B. W. *Icarus* **36**, 304-329 (1978).
5. Harris, D. L. in *The Solar System Vol. III. Planets and Satellites* (eds Kuiper, G. P. & Middlehurst, B. W.) 272-342 (University of Chicago Press, 1961).
6. Johnson, T. V. *Icarus* **14**, 94-111 (1971).
7. Johnson, T. V. & McCord, T. B. *Astrophys. J.* **169**, 589-593 (1971).
8. Pollack, J. B. *et al.* *Icarus* **36**, 271-303 (1978).
9. Cruikshank, D. P. *Icarus* (in the press).
10. Morrison, D., Cruikshank, D. P. & Burns, J. A. in *Planetary Satellites* (ed. Burns, J. A.) 3-17 (University of Arizona Press, 1977).

11. Johnson, T. V. & Pilcher, C. B. in *Planetary Satellites* (ed. Burns, J. A.) 232-268 (University of Arizona Press, 1977).
12. Morrison, D. & Morrison, N. D. in *Planetary Satellites* (ed. Burns, J. A.) 363-378 (University of Arizona Press, 1977).
13. Caldwell, J. *Icarus* **25**, 384-396 (1975).
14. Matson, D. L., Johnson, T. V. & Fanale, F. P. *Astrophys. J. Lett.* **192**, L43-L46 (1974).
15. Herzberg, G. *Infrared and Raman Spectroscopy of Polyatomic Molecules* (Van Nostrand, New York, 1945).
16. Nakamoto, K. *Infrared Spectra of Inorganic and Coordination Compounds* (Wiley, New York, 1963).
17. Eischen, R. & Pliskin, W. *Advances in Catalysis* Vol. 10, 1-56 (Academic, New York, 1958).
18. Lebofsky, L. *Icarus* **25**, 205-217 (1975).
19. Nash, D. B. *J. geophys. Res.* **71**, 2517-2534 (1966).

Spectral evidence for SO_2 frost or adsorbate on Io's surface

William D. Smythe*†, Robert M. Nelson*, & Douglas B. Nash*

*Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91103

†Institute of Geophysics and Planetary Physics, University of California at Los Angeles, Los Angeles, California 90024

The infrared reflection spectrum of Io matches that from an SO_2 frost grown in the laboratory.

It has been recently suggested that adsorbed SO_2 and/or H_2S on the surface of a spectroscopically benign substrate can explain the IR spectrum of Jupiter's satellite, Io¹. SO_2 gas has been identified as a major constituent of the volcanic gas plumes recently discovered by the Voyager 1 spacecraft². We have measured the reflection spectrum of SO_2 frost grown under low pressure at 140 K in the laboratory and find that it provides an excellent match to Io's spectrum in the 1.0-4.5 μm range.

The apparatus used and experimental methods have been described by Kieffer³ and Smythe⁴ respectively. Figure 1 compares the spectral geometric albedo of Io^{5,6} and our laboratory spectrum. Clearly SO_2 frost can explain the Io absorption bands at 1.35, 2.55, 2.80, 3.80, 3.90, 3.95, and 4.08 μm . Other candidate materials such as sulphates, nitrates and carbonates⁵ do not provide such a unique match to these bands.

Although SO_2 frost may be unstable on Io, except at the poles or nightside (see ref. 7), adsorbed SO_2 should be present on the surface particulates due to the extensive volcanic outgassing. The absorption bands of adsorbed and condensed species are known to be similar in wavelength⁸. Slight shifts in band positions and depths may reflect areal or temporal variation in frost/adsorbate or condensate/substrate contribution ratios to a multicomponent spectrum. Mixtures of alkali sulphides, free sulphur and adsorbates of SO_2 or H_2S have been suggested to explain Io's overall UV-VIS-IR reflectance¹. We conclude here that SO_2 as a frost or an adsorbate is the primary contributor to Io's IR spectral reflectance.

We thank Dennis L. Matson and Robert Carlson for helpful conversations, Hugh Kieffer for use of his laboratory at UCLA, and Larry Pleskot for assistance. We thank Dale Cruikshank for a revised Io spectra. This represents the results of one phase of research carried out at the JPL under NASA grant NAS 7-100. R.M.N. is an NRC resident research associate.

Received 8 June; accepted 17 July 1979.

1. Nash, D. B. & Nelson, R. M. *Nature* **280**, 763-766 (1979).
2. Hanel, R. *et al.* Press Release (American Geophysical Union meeting, Washington, D.C., 5 May 1979).
3. Kieffer, H. H. *Appl. Opt.* **8**, 2497-2500 (1969).
4. Smythe, W. D. thesis, Univ. California, Los Angeles (1979).
5. Pollack, J. B. *et al.* *Icarus* **36**, 271-303 (1978).
6. Cruikshank, D. P. *Icarus* (in the press).
7. Lebofsky, L. A. *Icarus* **25**, 205-217 (1975).
8. Eischen, R. P. & Pliskin, W. A. *Adv. Catalysis* **10**, 1-56 (1958).

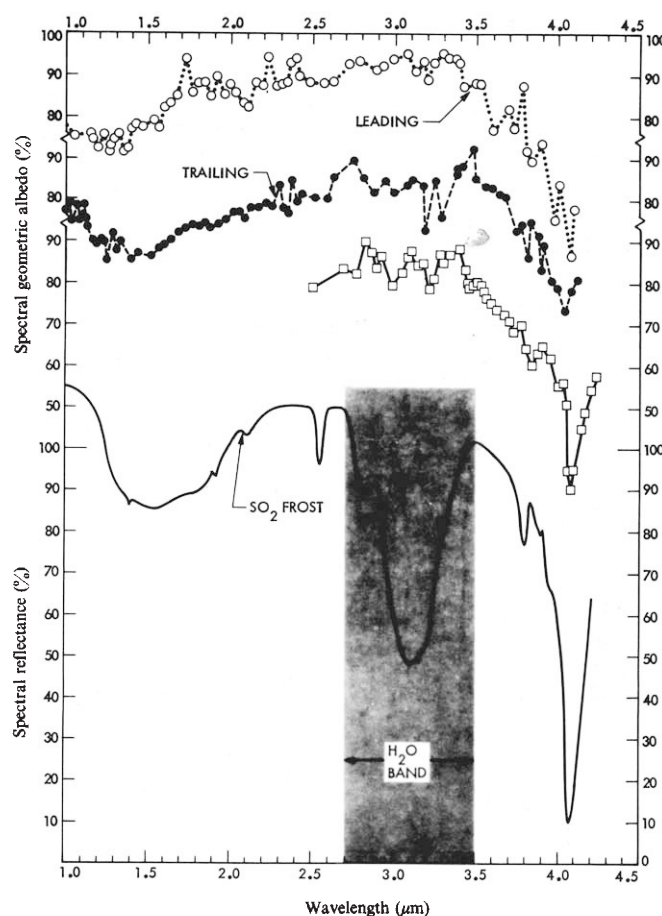


Fig. 1 *a*, Spectral geometric albedo of Jupiter's satellite Io, 1.0-4.5 μm based on previous studies^{5,6}, normalised as in ref. 1. $\circ$, $\bullet$, Data from ref. 5; $\square$, data from ref. 6. *b*, Spectral reflectance of SO_2 frost grown in the laboratory. The 3.1 μm H_2O band (stippled area) is due to contaminant water.

Auroral hiss observed near the Io plasma torus

D. A. Gurnett & W. S. Kurth

Department of Physics and Astronomy, The University of Iowa, Iowa City, Iowa 52242

F. L. Scarf

TRW Defense and Space Systems Group, 1 Space Park, Redondo Beach, California 90278

A band of whistler-mode noise identified as auroral hiss has been observed on the inner edge of the Io plasma torus. This noise provides evidence for the existence of aurora-like charged particle beams on magnetic field lines through the inner edge of the torus. These beams probably consist of low-energy electrons and may be associated with field-aligned currents linking the plasma torus to the jovian ionosphere.

DURING the Voyager 1 flyby of Jupiter on 5 March 1979 an unusual band of electric field noise was detected by the plasma wave instrument near the Io plasma torus. This band of noise has spectral characteristics very similar to a type of whistler-mode emission called 'auroral hiss'⁴ which is commonly observed in the Earth's auroral regions¹⁻⁵. Auroral hiss is believed to be produced by low-energy (10 eV to 1 keV) beams of auroral electrons by a Cerenkov radiation mechanism^{6,7}. Because the presence of auroral hiss near the Io torus may indicate the occurrence of charged particle acceleration processes similar to the Earth's auroral zones, we have undertaken a detailed analysis of these noise bands. This article describes the characteristics of the auroral hiss emissions detected at Jupiter, compares these characteristics with similar emissions detected in the Earth's magnetosphere, and comments on the implications regarding the interaction of the plasma torus with the jovian magnetosphere. The plasma wave instrumentation on Voyager 1 and an initial survey of the Voyager 1 plasma wave observations at Jupiter are given elsewhere^{8,9}.

Observations

The noise bands of interest were observed near the inner edge of the Io plasma torus on both the inbound and outbound portions of the Voyager 1 pass by Jupiter. For a discussion of the plasma torus formed by material escaping from Jupiter's satellite Io see refs 10 and 11. The plasma wave electric field intensities detected near the closest approach to Jupiter are shown in Fig. 1b. A sketch of the spacecraft trajectory in relation to the plasma torus is shown in Fig. 1a together with the electron plasma frequency, f_p^- , obtained from the planetary radio astronomy experiment¹¹ and the electron gyrofrequency, f_g^- , obtained from the magnetometer experiment¹².

The emissions identified as auroral hiss are indicated by the cross-hatched areas in Fig. 1b. The distinguishing characteristic of these emissions is the systematic change in the emission frequency with spatial position, increasing with increasing time for the event at 09.30 UT, and decreasing with increasing time for the event at 14.00 UT. Because of the limited frequency-time resolution of the intensity measurements in Fig. 1 it is not possible to identify these emissions as auroral hiss on the basis of these data alone. Fortunately, for the event at 09.30 UT, a series of wideband waveform measurements was available, providing much better frequency-time resolution. The wideband waveform mode of operation uses the 115 kbit s⁻¹ telemetry system normally used for pictures to transmit a 12 kHz bandwidth of electric field waveforms to the ground. Because the 115 kbit s⁻¹ telemetry link must be time shared with the imaging system, the waveform measurements are not obtained continuously. During

the period around 09.30 UT, one 48-s burst of waveform data was being obtained every 192 s. No wideband measurements were obtained for the period around 14.00 UT. Three consecutive 48-s spectrograms of the wideband waveform measurements are shown in Fig. 2 for the event at 09.30 UT. These spectrograms show a broadband emission with a very sharp low frequency cutoff which increases linearly with increasing time. The broadband character of the emission and the smoothly varying low frequency cutoff provide an almost unmistakable identification of this emission as a V-shaped auroral hiss event of the type described by Gurnett², McEwen and Barrington³ and others. As will be discussed, the low frequency cutoff is a whistler-mode propagation effect. At the Earth the spatial variation of the cutoff frequency produces a characteristic V shape on a frequency-time spectrogram with the centre of the V located in a region of intense low-energy electron precipitation⁴.

The auroral hiss emissions detected by Voyager are unusual in that only one branch of the V-shaped low frequency cutoff is evident. Such unsymmetrical V-shaped auroral hiss events are also occasionally seen in the Earth's auroral zone, usually in association with a large density gradient. A similar relationship is evident for the events detected by Voyager. As Fig. 1 shows both of the auroral hiss events occur near a very large density gradient at the inner edge of the plasma torus. Whether the missing branch of the V is a propagation effect related to the density gradient or is simply obscured by the more intense chorus and hiss emissions⁹ inside the plasma torus is not known.

Analysis

For auroral hiss at the Earth it is generally believed^{7,13} that the V-shaped low-frequency cutoff is a spatial propagation effect which occurs for short wavelength waves near the resonance cone. As will be shown the beamwidth of whistler-mode radiation near the resonance cone is frequency dependent, with the highest frequencies having the largest beamwidth. As the spacecraft approaches the source the highest frequencies are detected first, because these ray paths can propagate at the largest angle to the magnetic field. Progressively lower frequencies are detected as the spacecraft comes closer to the source.

The spatial dependence of the cutoff frequency is best analysed using the index of refraction construction shown in Fig. 3b. The ray path at a given wave normal direction is perpendicular to the index of refraction surface at that wave normal angle¹⁴. Figure 3b shows that the angle between the limiting ray path direction and the magnetic field, $\psi_{\max}$, is given by $\psi_{\max} = \pi/2 - \theta_{\text{Res}}$, where θ_{Res} is the resonance cone angle. Using Stix's notation¹⁴, the angle $\psi_{\max}$ is given by:

$$\cot^2 \psi_{\max} = \tan^2 \theta_{\text{Res}} = -\left(\frac{P}{S}\right) \quad (1)$$

All waves near the resonance cone will have ray directions less than this limiting angle. For the conditions applicable to the Voyager 1 observations, $f_g^- \ll f_p^-$, and $f_g^+ \ll f < f_g^-$, the terms S and P can be approximated by

$$P = -\left(\frac{f_p^-}{f}\right)^2 \quad \text{and} \quad S = \left(\frac{f_p^-}{f_g^+}\right)^2 \left[1 - \frac{f_{\text{LHR}}^2}{f^2}\right] \quad (2)$$

where $f_{\text{LHR}} = (f_g^+ f_p^-)^{1/2}$ is the lower-hybrid resonance frequency.

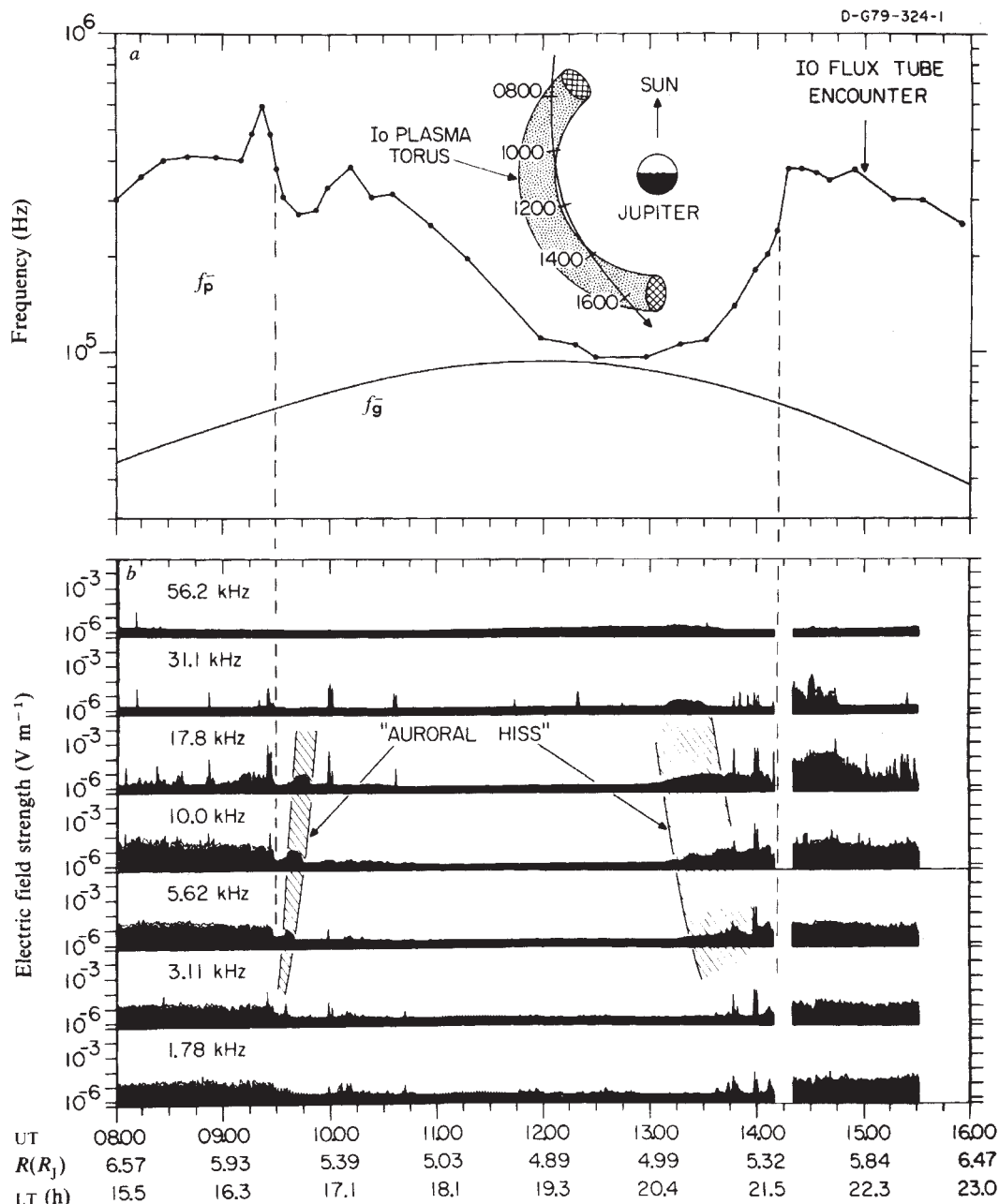


Fig. 1 The plasma wave electric field measurements for the Voyager 1 pass through the Io plasma torus. The emissions identified as whistler-mode auroral hiss are identified by the cross-hatched areas. In both cases the auroral hiss emissions were observed near an abrupt gradient in the plasma density (see the vertical dashed line) near the inner edge of the torus.

Combining these terms gives the following equation for the limiting ray path angle as a function of frequency,

$$\tan^2 \psi_{\max} = \frac{f^2 - f_{LHR}^2}{(f_g)^2} \quad (3)$$

Equation (3) shows that the limiting ray path angle increases monotonically with increasing frequency for $f^2 \gg f_{LHR}^2$, and goes to zero at $f = f_{LHR}$. If we now consider an idealised point source emitting all frequencies the resulting radiation will illuminate a conical region along the magnetic field with a beamwidth which increases with increasing frequency and goes to zero at f_{LHR} . This frequency-dependent beamwidth is illustrated in Fig. 3a. A spacecraft passing through the illumination region first detects the highest frequency, f_2 , then progressively lower frequencies, f_1 , until the lowest frequency f_{LHR} occurs when the spacecraft is on the magnetic field line, through the source. The frequency-dependent beamwidth explains the V-shaped spatial variation of the low frequency cutoff.

For the Voyager auroral hiss observations, as for the terrestrial auroral hiss, it seems more appropriate to consider a line source extending azimuthally along the torus rather than a point source. This configuration is supported by the fact that somewhat similar symmetrical auroral hiss emissions were observed

on both the inbound and outbound crossings at substantially different local times. Ignoring the curvature of the magnetic field lines, the ray path problem can be reduced to a simplified two-dimensional geometry by introducing a coordinate x , which is the perpendicular distance from the spacecraft to the magnetic field line through the source, and a distance h , which is the height of the spacecraft above the source (see Fig. 3). Simple trigonometry shows that for the limiting ray path $x/h = \tan \psi_{\max}$. Substituting into equation (3) and simplifying gives the following equation for the cutoff frequency as a function of x ,

$$f^2 = f_{LHR}^2 + \left(\frac{x}{h}\right)^2 (f_g)^2 \quad (4)$$

This equation is a hyperbola. Even when the magnetic field curvature is considered, equation (4) is still valid to second order provided x is measured perpendicular from the magnetic field through the source.

To obtain the approximate location of the source, equation (4) has been adjusted to fit the measured cutoff frequencies. The cutoff frequencies, which can be measured with very good accuracy (~ 100 Hz), are shown by dots in Fig. 3c along with the best fit hyperbola representing the propagation cutoff. Only three free parameters occur in equation (4), the lower-hybrid

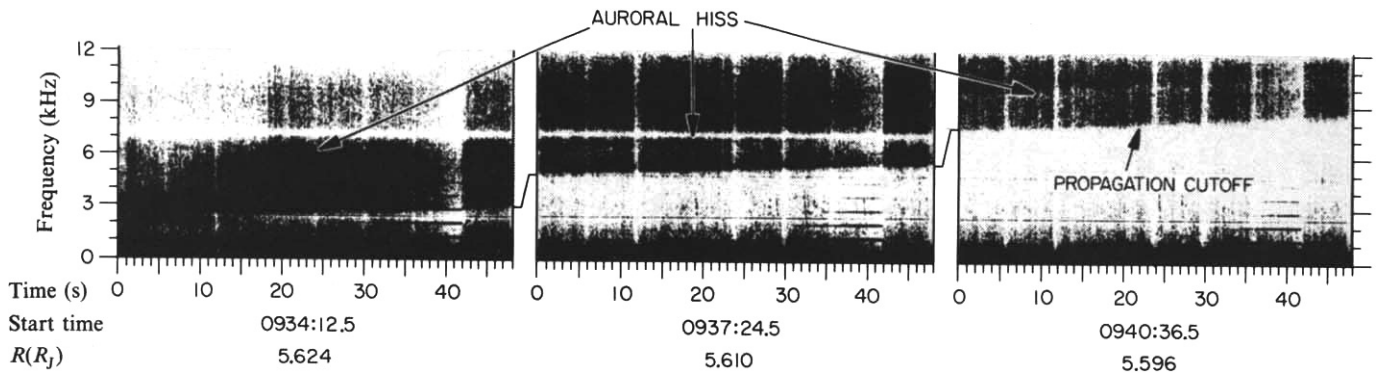


Fig. 2 A sequence of wideband spectrograms for the event near 09.30 UT in Fig. 1. The broadband character of the emission and the sharp low frequency propagation cutoff increasing linearly with time provide a clear identification of this emission as auroral hiss.

resonance frequency, the location of the magnetic field line through the source, and the height h of the spacecraft above the source. In principle, the lower-hybrid resonance frequency can be determined from the plasma parameters and need not be considered a free parameter. However, the lower-hybrid resonance frequency depends on the ion composition which varies considerably. As the ion composition is not well known along the ray path, f_{LHR} is considered an unknown parameter. The best fit parameters indicate that the magnetic field line through the source was crossed at 09.32 UT, that the distance h to the source was $0.65 R_J$, and that the lower-hybrid resonance frequency was 1.56 kHz. At the time of the closest approach to the source, Voyager was $0.51 R_J$ north of the magnetic equator. Two possible source positions must be considered, either northward along the magnetic field line towards Jupiter, or southward along the magnetic field towards the magnetic equator. The northward position would place the source approxi-

mately $0.65 R_J + 0.51 R_J = 1.16 R_J$ north of the equator on the $L = 5.6$ field line. The southward position would place the source approximately $0.65 R_J - 0.51 R_J = 0.14 R_J$ south of the magnetic equator. These two source positions are illustrated in Fig. 4a and b. Since the plasma wave instrument does not have any capability for determining the direction of propagation, it is not possible to distinguish experimentally these two source locations.

Although the north-south ambiguity cannot be resolved, certain indirect arguments can be made favouring each of the two source positions shown in Fig. 4. The equatorward source is favoured by the fact that this location, only $0.14 R_J$ south of the magnetic equator, coincides with an obvious symmetry point where a source could reasonably be expected. The best fit lower-hybrid resonance frequency is not, however, in good agreement with the lower-hybrid resonance frequency expected in this region. Since the inner region of the Io torus is thought to

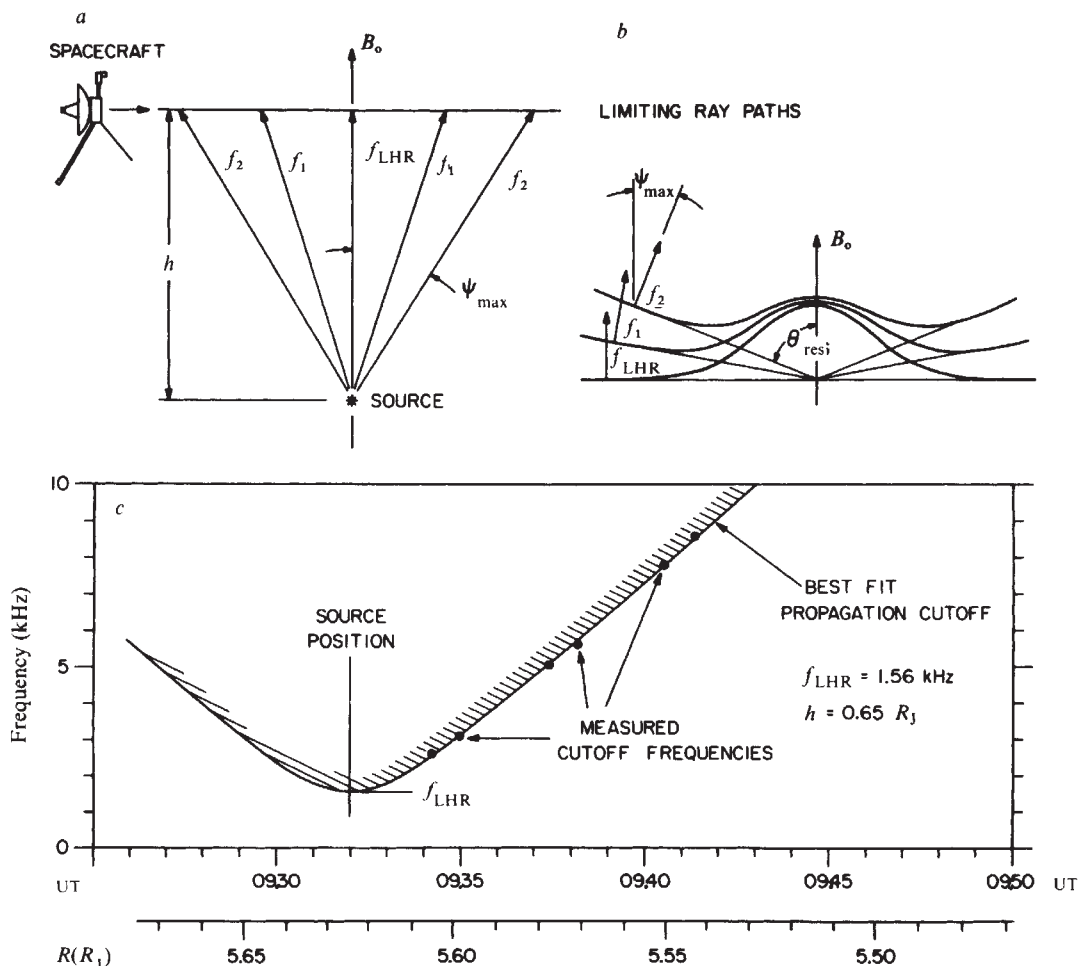


Fig. 3 a and b show that the beamwidth of whistler-mode emissions near the resonance cone increases with increasing frequency. A spacecraft approaching a point source, therefore, detects the highest frequencies first, forming a V-shaped low frequency cutoff. c, The best fit propagation cutoff for the auroral hiss emission near 09.30 UT. Only one branch of the V-shaped cutoff was observed.

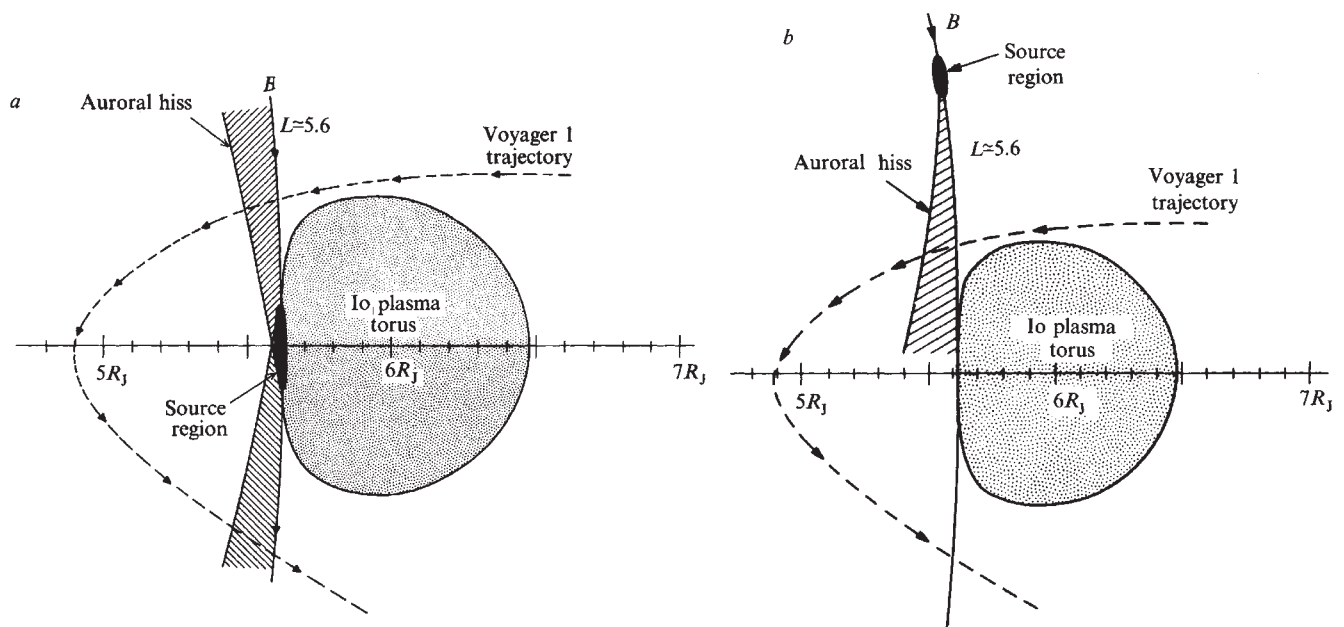


Fig. 4 The two possible source positions obtained from the best fit analysis in Fig. 3. *a*, The equatorward source position. *b*, The northward source position which has a very close analogy to auroral hiss emissions at the Earth, with the Io plasma torus playing a role comparable to the ionosphere at the Earth. The auroral hiss emissions strongly suggest the occurrence of low-energy electron beams and field line currents associated with the large density gradient near the inner edge of the plasma torus.

be dominated by heavy ions^{10,15}, the lower-hybrid resonance frequency is expected to be of the order of 300–500 Hz inside the torus. The best fit lower-hybrid resonance frequency, 1.56 kHz, seems to be more consistent with a plasma in which protons are the dominant ion species. As protons are expected to be a more important constituent outside the torus, this fact would tend to favour a source well away from the equator, such as in Fig. 4*b*. Note, however, that the lower-hybrid resonance frequency has the largest uncertainty of all the free parameters in equation (4). This difficulty occurs mainly because no wide-band data is available around the time, 09.32 UT, when the cutoff frequency is near the apex of the hyperbola. The uncertainty in f_{LHR} , together with the absence of suitable information on the ion composition away from the torus, makes it very difficult to distinguish between the models in Fig. 4 on the basis of the lower-hybrid cutoff frequency.

Discussion

These observations of auroral hiss provide strong evidence of auroral-like charged particle beams originating from within, or near, the Io torus. As the intensity and spectral characteristics of the auroral hiss detected at Jupiter are very similar to the auroral hiss observed at the Earth, it is expected that the radiation should be generated by comparable particle intensities, namely electrons with energies from about 10 eV to 1 keV and fluxes in the range 10^8 to 10^{10} electrons cm^{-2} . The existence of comparable low-energy electron precipitation from the torus is already implied by the UV spectrometer observations of aurora at the foot of the magnetic field lines through the torus. The fact that the auroral hiss is observed over a very small spatial region with very sharply defined cutoff frequencies implies that the electron beam generating the radiation must be very narrow, not extending over a range of L -values more than about 0.02. The source position of the auroral hiss shows that

the electron beam occurs on an L -shell which coincides almost exactly with the abrupt decrease in plasma density at the inner edge of the plasma torus. Again in analogy with the terrestrial auroral zone, it seems likely that the auroral hiss and associated electron beams would be associated with a thin sheet of current which flows along the magnetic field lines linking the inner edge of the plasma torus with the polar ionosphere of Jupiter. At present we are unable to determine the direction of the current in this sheet, as this depends on the direction of propagation of the auroral hiss. For the model in Fig. 4*a*, the auroral hiss and the electrons producing this hiss would be moving away from the equatorial plane, which implies a current directed along the magnetic field lines from Jupiter towards the Io torus. For the model in Fig. 4*b*, the directions would be reversed.

At the present early stage of data analysis, no specific evidence is available, other than the auroral hiss observations, to confirm or disprove the existence of low energy electron beams and field-aligned currents at the inner edge of the Io plasma torus, although such investigations are under way (N. Ness and H. Bridge, personal communication). The likelihood of such field-aligned currents and particle precipitation effects is, however, considered quite high, as several processes can give rise to field-aligned currents at abrupt gradients in the plasma density. Given that a field-aligned current of sufficient magnitude exists, the acceleration of particles to large energies could proceed via the development of parallel electric fields, as in the Earth's auroral zones. Observations of auroral hiss at the Earth have in fact been suggested as being directly related to regions of parallel electric fields and auroral particle acceleration¹⁶.

Research at The University of Iowa was supported by NASA through contract 954013 with the Jet Propulsion Laboratory and through grant NGL-16-001-043 with NASA Headquarters. Research at TRW was supported by NASA through contract 954012 with the Jet Propulsion Laboratory.

Received 16 July; accepted 31 July 1979.

1. Helliwell, R. A. *Whistlers and Related Ionospheric Phenomena*, 288 (Stanford University Press, Palo Alto, 1965).
2. Gurnett, D. A. *J. geophys. Res.* **71**, 5599 (1966).
3. McEwen, D. J. & Barrington, R. E. *Can. J. Phys.* **45**, 13 (1976).
4. Gurnett, D. A. & Frank, L. A. *J. geophys. Res.* **77**, 172 (1972).
5. Hoffman, R. A. & Laaspere, T. *J. geophys. Res.* **77**, 640 (1972).
6. Maggs, J. E. *J. geophys. Res.* **81**, 1707 (1976).
7. James, H. G. *J. geophys. Res.* **81**, 501 (1976).

8. Scarf, F. L. & Gurnett, D. A. *Space Sci. Rev.* **21**, 289 (1977).
9. Scarf, F. L., Gurnett, D. A. & Kurth, W. S. *Science* **204**, 991 (1979).
10. Broadfoot, A. L. *et al. Science* **204**, 979–982 (1979).
11. Warwick, J. W. *et al. Science* **204**, 995–998 (1979).
12. Ness, *et al. Science* **204**, 982–987 (1979).
13. Mosier, S. R. & Gurnett, D. A. *J. geophys. Res.* **74**, 5675 (1969).
14. Stix, *The Theory of Plasma Waves* (MacGraw-Hill, New York, 1962).
15. Bridge, H. S. *et al. Science* **204**, 987–991 (1979).
16. Shawhan, S. D. *Solar System Plasma Physics*, Vol. 3 (eds Lanzerotti, L. J., Kennel, C. F. & Parker, E. N.) 793 (North-Holland, Amsterdam, 1979).

Pre-Voyager velocities, accelerations and shrinkage rates of jovian cloud features

Reta F. Beebe & Lou Ann Youngblood

Department of Astronomy, New Mexico State University, Las Cruces, New Mexico 89003

The Voyager 1 data yield detailed information on the latitudinal dependence of velocities in the jovian atmosphere. This enhances the possibility of interpreting the time-dependent behaviour of long-lived jovian cloud features. This article summarises pre-Voyager velocity data, stressing gross differences with the Voyager 1 results. Analytic expressions are given for average drift rates and shrinkage rates for the Great Red Spot and white ovals (located at -23° and -34° latitude, respectively) from 1943 to 1979.

HISTORICAL records of longitudinal positions and size of jovian cloud features include visual transit measurements by Reese from 1943 to 1964^{1,2} dovetailed with photographic observations obtained at New Mexico State University (NMSU) from 1963 to the present³⁻⁹. Because Reese assisted in the development of the measuring technique for the photographic data while continuing to make visual transit measurements, and then continued with the photographic programme through 1976, these data should be self-consistent. Although the visual transits have a large measuring error (standard error, 1–2 degrees), the average values of size and longitudinal position of the centre of each feature near the time of opposition can be used to derive long-term shrinkage rates and average drift rates.

Since 1974, a concentrated effort has been directed towards obtaining longitudinal and latitudinal position measurements of all jovian cloud features which can be consistently identified over periods of a few weeks or more. Figure 1 presents a composite of the velocities obtained between April 1974 and May 1979, where the horizontal bar represents the range of velocities that were observed at a specified latitude. Latitudes are reported as planetographic² and velocities are expressed relative to System III longitude¹⁰.

This study provides a finer latitudinal discrimination than that presented by Smith and Hunt¹¹. The velocity at a particular latitude cannot always be determined because relatively large features must be present for groundbased detection; however, the verification of the presence of all of the previously measured latitudinal velocities plus the determination of the velocity patterns around the Great Red Spot (GRS) and ovals by Voyager 1 data¹² should provide a large contribution to under-

standing long-term cloud motions. Voyager data^{12,13} yield a velocity pattern that is more symmetric about the equator than the data shown in Fig. 1. Note that large features in the southern hemisphere tend to dominate the groundbased velocity profile, while the measured velocities in the northern hemisphere are obtained by tracking smaller features whose diameters are $\sim 5,000$ km and whose albedos contrast strongly with their surroundings. These small, strongly contrasting features are rare in the southern hemisphere at groundbased resolutions; hence, further interpretation of the historical data requires the more complete understanding of the atmospheric velocity structure now being obtained from the Voyager data.

The formation of the three white ovals occurred in 1938, following an obscuration of the south temperate belt² by a high-albedo cloud formation. Three dark regions appeared within the belt and gradually the intervening bright areas coalesced into ovals. During their formation they were hazy and diffuse; however, by 1953 they had contracted to 26,000 km in longitudinal extent, or length, and have continued to decrease to a present length of 12,000 km. In 1941, the ovals drifted at a rate of 7.5 m s^{-1} and have since slowed down to 4.0 m s^{-1} . Even

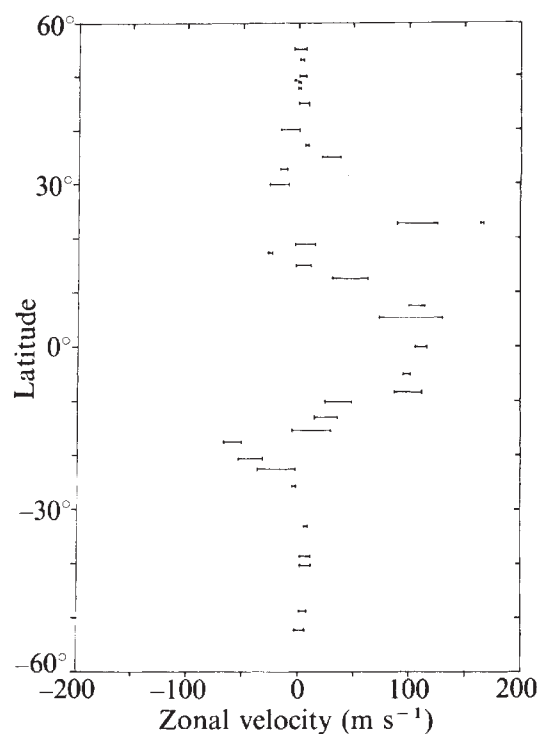


Fig. 1 Zonal velocities derived from groundbased photographic data. Positional measurements have been reduced to System III longitude¹⁰ and drift rates are obtained for features that can be identified from day to day on the photographs. An automated measuring program allows accurate determination of latitude, as well as measurement of a large number of features.

Table 1 Least squares drift rate where longitudinal position in System III equals $C_0 + C_1\Delta t + C_2\Delta t^2$

Solution	C_0 (deg)	C_1 (deg yr ⁻¹)	C_2 (deg yr ⁻²)	Acceleration (m s ⁻²)
GRS	86.16	102.4	-0.1124	1.310×10^{-10}
Forced fit of the three	-11.14	-148.5	-1.3448	-1.442×10^{-9}
white ovals	116.47			
	231.91			

Δt is the time (yr) elapsed since 1965.0.

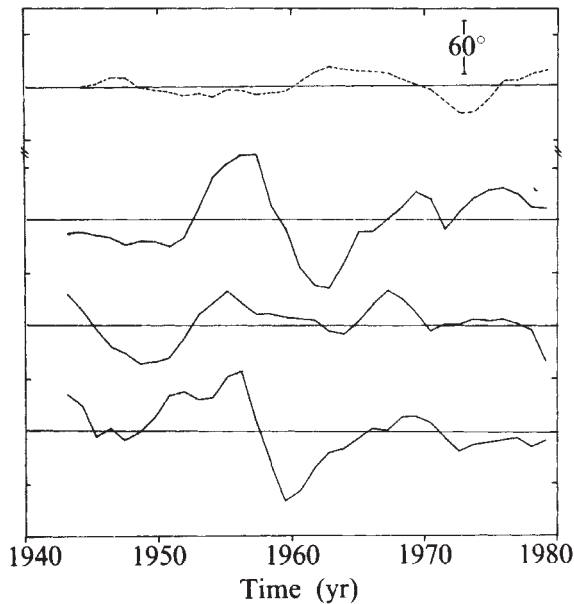


Fig. 2 Residuals of longitudinal positions. These residuals have been obtained assuming that the GRS (dotted line) drifts at constant acceleration throughout the time interval and that the three white ovals (solid line) move together as a system undergoing constant acceleration. The vertical scale is displacement in degrees of longitude with the magnitude of displacement indicated by the vertical bar at the upper left of the chart. Note that displacement below the line represents observed drift rates that are larger than the second-order polynomial fit implies.

though ovals drifting at a rate of 3 m s^{-1} were observed at this latitude from 1915 to 1938¹⁴, there are no recorded observations of their disappearance. These earlier ovals were not observed to dissipate but, instead, became obscured by the general brightening of the south temperate belt in 1938.

Historical records indicate that the GRS has also decreased in length. During the period 1879–82, its dimensions were 40,000 by 13,000 km (ref. 2). Its length has shrunk to 25,000 km at present, while its width has remained approximately 13,000 km (ref. 9). For the past 40 yr there has been only minor observable activity in the south tropical zone, which lies between the white ovals and the GRS, and the drift rate of the GRS has been approximately constant.

To evaluate the rate of acceleration of these features quantitatively, a data set consisting of the longitude of the centre of the feature at the time of opposition for each apparition since 1943 was subjected to a least squares procedure which fit the data to a second-order polynomial. In the case of the three white ovals, the solution required that data from the three ovals have the same time dependence. The coefficients of the second-order fit are shown in Table 1 along with the acceleration of the features, assuming constant acceleration relative to System III longitude throughout the time interval. The residuals of the fit are plotted as a function of time in Fig. 2.

The GRS rotates more slowly than System III longitude (towards increasing longitude), and the white ovals rotate more rapidly; therefore, the coefficients in Table 1 suggest that the drift rates at both latitudes tend towards a common value. A disturbance in the south tropical zone was observed to catch up with and pass the GRS on 23 December 1970, accelerating the GRS as it did so⁵. Dark streaks were also observed in the south tropical zone in 1946–47 and 1955–57 (Reese, personal com-

munication). Inspection of the GRS residuals suggests that the GRS is predictably accelerated by this activity, while the ovals are insensitive to it.

Because the GRS is larger than the ovals, Fig. 2 suggests that the smaller the oval cloud feature the more variable its velocity. A contradiction of this assumption is that observed velocities of small white ovals at -41° latitude are quite constant at 7 m s^{-1} and the spacing between these individual ovals does not change as it does for the larger white ovals at -34° .

The rate of decrease in observable areas of the GRS and white ovals is proportional to the length of the features. The width of these four features is apparently determined by the surrounding zonal flow pattern; thus, their width has varied little while their length has decreased as a function of time. Shrinkage rates have followed two patterns. In the case of the GRS, the shrinkage was approximately linear throughout the interval from 1943 to 1979. The white ovals decreased rapidly in length between 1943 and 1966; however, since 1967 the shrinkage rate has been

Table 2 Rates of shrinkage of the length of the GRS and white ovals

Case	Feature	Time interval	Formulae	L_0	C	s.e.
1	GRS	1944–79	$L = L_0 + C \times (t - 1944.115)$	25.9	-0.114	2.2
2	White ovals	1944–79	$C \ln L = \ln L_0 + (t - 1944.115)^{1/2}$	68.7 75.6 64.6	-0.324	0.111*
3	White ovals	1944–66	$\ln L = \ln L_0 + C(t - 1944.15)^{1/2}$	67.1 77.1 62.3	-0.317	0.113*
4	White ovals	1967–79	$L = L_0 + C \times (t - 1967.005)$	13.7 13.9 13.3	-0.249	0.98

L is length of feature in degrees (to convert to kilometres multiply L by 1,160 km per deg of longitude for the GRS and 1,060 km per deg for the white ovals). t is date (yr). s.e. is the standard error.

* These errors correspond to $\Delta \log L$.

approximately linear. The data for the white ovals indicate that the shrinkage rate during the interval 1943–55 was slower than that predicted by an exponential function; hence, the function $\exp[a(t-t_0)^{1/2}]$ has been used to generate the analytical expressions given in Table 2. The residuals from cases 3 and 4 do not indicate a more convincing fit than case 2; however, observational scatter and the differences in the sizes of the three ovals during their early states of contraction do not recommend additional analysis. Because case 4 is less influenced by early variations in the length of the individual ovals, and because they have been virtually indistinguishable in size since 1948, it is more reliable for predicting future contraction of the ovals. Case 1 predicts that the GRS will span 21° in 1985 (24,500 km compared with the present value of 25,000 km) and, according to case 4, the white ovals will extend $9.1^\circ \pm 0.3$ in 1985 (9,700 km compared with 12,000 km, currently); hence, shrinkage rates are small for the GRS, but are significant for the white ovals.

Conclusions

Our data provide smoothed values for shrinkage rates and acceleration of the GRS and white ovals for the period from 1943 to 1979. Additional analysis of data from 1963 to the present will be presented later.

We thank E. Reese, D. Cress, T. Bruce, S. Murrell and C. Knuckles for assistance. Funding was from NASA grant NGL 32-003-001 and NSF grant AST-7722962.

Received 16 July; accepted 31 July 1979.

- Alexander, A. F. *J. Br. astr. Ass.* **63**, 251–334 (1953).
- Peek, B. M. *The Planet Jupiter* (Faber, London, 1958).
- Reese, E. J. *Icarus* **12**, 249–257 (1970).
- Reese, E. J. *Icarus* **14**, 343–354 (1971).
- Reese, E. J. *Icarus* **17**, 57–72 (1972).
- Beebe, R. F. & Reese, E. J. *NMSU Obs. TN-76-50* (1976).

- Beebe, R. F. *et al.* *NMSU Obs. TN-77-58* (1977).
- Beebe, R. F. *et al.* *NMSU Obs. TN-78-61* (1978).
- Beebe, R. F. *et al.* *NMSU Obs. TN-79-63* (1979).
- Riddle, A. S. & Warwick, J. W. *Icarus* **27**, 457–459 (1976).
- Smith, B. A. & Hunt, G. E. in *Jupiter* (ed. Gehrels, T.) 564–585 (University of Arizona Press, 1976).
- Ingersoll, A. P. *et al.* *Nature* **280**, 773–775 (1979).
- Mitchell, J. L. *et al.* *Nature* **280**, 776–778 (1979).
- Reese, E. J. *Icarus* **17**, 704–706 (1972).

Zonal velocity and texture in the jovian atmosphere inferred from Voyager images

A. P. Ingersoll*, R. F. Beebe†, S. A. Collins‡, G. E. Hunt§, J. L. Mitchell‡, P. Muller§, B. A. Smith|| & R. J. Terrile‡

* Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena 91125

† Department of Astronomy, New Mexico State University, Las Cruces 88003

‡ Jet Propulsion Laboratory, Pasadena, California 91103

§ University College London, London WC1, UK

|| Department of Planetary Sciences, University of Arizona, Tucson 85721

Measurements of the latitudinal profile of zonal velocity are used to estimate the absolute vorticity gradient. A classification scheme of small features visible at resolutions of 100 km or better is presented

THE first report¹ of the Voyager imaging science team following the 5 March 1979 encounter described Jupiter's changing appearance at resolutions down to 10 km, over intervals as small as 1 h. Examples of small-scale convection, rapid variations of features, and complex interactions of closed vortices were presented. This article extends these results in two ways. First, we present measurements of the latitudinal profile of zonal (eastward) velocity, from which we estimate the absolute vorticity gradient. Second, we present a classification scheme based on texture, by which we mean the patterns of small features visible at resolutions of 100 km or better. In accompanying articles we discuss characteristic structures and flow features of separated regions², and equatorial wave phenomena as observed by Voyager³.

To date, we have not used the full potential of multi-colour imaging coupled with IR observations to infer vertical structure. Below we assume that all observed motion takes place at the same pressure altitude. We have also not fully analysed the high resolution (10–100 km) frames taken during the last few days before encounter. Thus we also assume that the frames that have been analysed are representative of the planet. Finally, we have not yet made detailed comparisons between Voyager data and results of theoretical models. Such comparisons will be made as the empirical basis of understanding matures.

Zonal velocity structure

Figure 1 shows zonal velocities with respect to Jupiter's interior⁴ computed from displacements of cloud features that are about 100–200 km in size over time intervals of 10–30 h (1–3 jovian rotations) in late February 1979. We find no systematic differences between these results and those measured from frames taken 2 h apart, although velocities in the latter case are more uncertain. Figure 1 shows the work of several individuals analysing different pairs of frames. The spread of points provides an estimate of our measurement error.

The latitudes in Fig. 1 and elsewhere in this issue are planetographic, with assumed equatorial and polar radii of 71,400 and 66,770 km, respectively. Latitudes in the first Voyager report¹ were planetocentric, but were erroneously reported as planetographic. By definition, planetocentric and planetographic latitudes are the elevation angles with respect to the equatorial plane of a line from the centre of the planet and a line perpendicular to the surface of the planet, respectively.

Comparing these Voyager results (Fig. 1) with 100 yr of Earth-based results^{5–8}, there is remarkable constancy in latitude of the major zonal currents. Virtually every major current that has ever been seen in 100 yr of Earth-based telescopic observations is detectable in one 10-h sequence of Voyager observations. This is especially striking when one recalls that many major currents (for example at 23°N, 29°S and 44°S) have gone undetected for years at a time⁶. We conclude, therefore, that the currents are quasi-permanent, but a lack of suitable markings sometimes renders them invisible from Earth.

The constancy in latitude of the major currents contrasts with the changing visual appearance (colour and albedo) of Jupiter. For example, during the 4.25 yr between the Pioneer 11 encounter⁹ and the Voyager 1 encounter^{1,10} the bright white zone initially at 12°–24°S became narrower by a factor of 2, and the zone currently at 18–30°N became wider by a similar factor.

Compared to the visual appearance of Jupiter, the zonal velocity exhibits very little north–south (hemispheric) asymmetry. In the north, eastward velocities tend to be stronger, and westward velocities weaker than in the south. The local velocity minimum near the equator apparently lies at 1°S (ref. 6). But the velocity profile in Fig. 1 exhibits greater symmetry about the equator than a map based on structural features (see below and ref. 2) or a colour image of the planet¹.

The maximum speeds of the zonal jets measured by Voyager tend to be greater than those measured from Earth^{5–8}. Earth-based observers tend to measure the larger, longer-lived spots, which generally lie between the major currents and do not participate in the high-speed flow. The Earth-based data shown in Fig. 1 are from a very intensive study⁸ for the period 1974–79, and show much more of the high speed motion than earlier Earth-based studies^{5–7}.

The data of Fig. 1 may be used to compute the latitudinal gradient $d\zeta_a/dy$ of absolute vorticity associated with the zonal wind profile $\bar{u}(y)$. By definition,

$$\frac{d\zeta_a}{dy} = \beta - \bar{u}'' = \frac{2\Omega \cos \theta}{r} - \frac{d^2\bar{u}}{dy^2}$$

where y is the northward coordinate, ζ_a is the vertical component of vorticity in inertial coordinates, Ω is the planetary rotation rate, r is the radius, θ is the latitude, and $\bar{u}''$ is the curvature of the velocity profile.

Since β is positive everywhere, the latitudinal average of $\beta - \bar{u}''$ is positive. The latitudes where $\beta - \bar{u}''$ is most likely to be negative are those where $\bar{u}''$ is large and positive. These are the latitudes of the westward jets. Accordingly, Fig. 2 shows curves of $\bar{u}$ for which $\bar{u}'' = \beta$, and these curves are centred on the westward jets and on the equator. Note that the measured points near each westward jet lie to the right of the smooth curve, that is

the observed velocity profile exhibits more curvature than β . We conclude that $\beta - \bar{u}''$ and hence $d\zeta_a/dy$ changes from positive to negative in the vicinity of the westward jets.

These Voyager data extend an earlier study¹¹ of Earth-based data which concluded that $\beta - \bar{u}'' \geq 0$ everywhere, with $\beta = \bar{u}''$ at the latitudes of the westward jets. The difference is a result of the tendency of Earth-based observers to underestimate the zonal speeds.

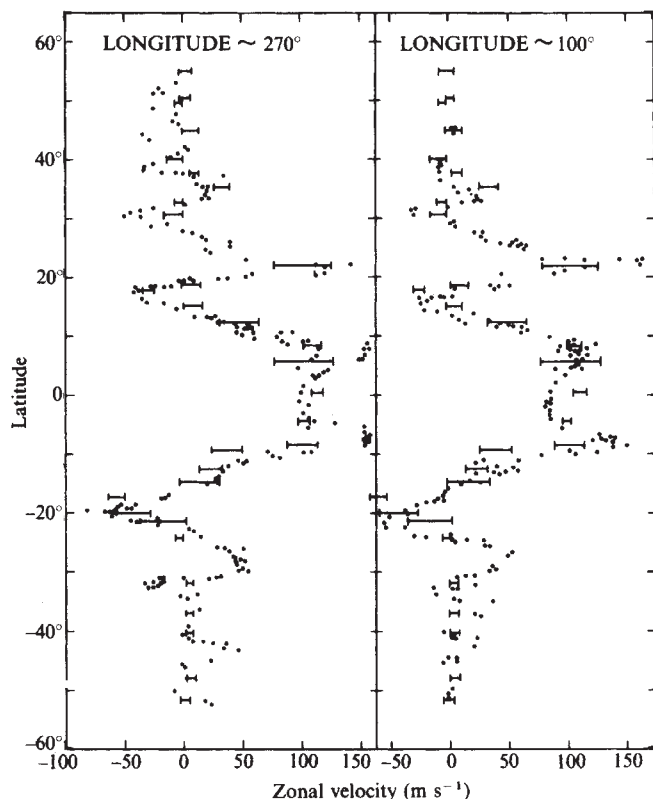


Fig. 1 Zonal velocities derived from Voyager 1 imaging data. Each point is computed from the displacement of a ~ 200 km cloud feature in frames 10–30 h apart. Earth-based measurements⁸ from the period 1974–79 are shown as horizontal bars. Because of frequent lack of suitable markings, not all currents are visible from Earth at the same time or during the same year. The spread of points and the length of horizontal bars are indicators both of measurement error and of variability on Jupiter. Two significant features of the zonal velocity profile are its constancy over periods of many years and its relatively high symmetry with respect to the equator.

According to the simplest two-dimensional models, the fact that $d\zeta_a/dy$ changes sign near the westward jets implies that the profile $\bar{u}(y)$ is unstable at these latitudes¹¹. The westward jets, particularly on their equatorward sides, are certainly more active and disturbed than the eastward jets^{2,12}. However, this interpretation depends on neglecting the vertical structure of the flow. The condition $\beta - \bar{u}'' \geq 0$ everywhere is the barotropic stability criterion¹³, a sufficient condition for stability of a zonal flow $\bar{u}(y)$. When $\bar{u}$ varies also with depth, this criterion is relevant provided the static stability (thermal stratification) is large enough. The condition $\beta - \bar{u}'' \geq 0$ is also the statistical equilibrium state of planetary turbulence, according to recent numerical studies^{14–16}. In these experiments, kinetic energy is introduced as small-scale eddies. Turbulent processes transfer this energy to a predominantly zonal flow whose latitudinal scale, defined by $(\bar{u}/\bar{u}'')^{1/2}$, increases until the condition $\bar{u}'' \approx \beta$ is reached. All of the above results depend on the model of vertical structure and lower boundary conditions^{17–18}, about which little is known. The observation that $\beta - \bar{u}''$ changes sign may provide

useful constraints on these unknown parameters of Jupiter's atmosphere.

Figure 2 also indicates that $\beta - \bar{u}'' > 0$ at the equator, since the measured points lie to the left of the smooth curve $\bar{u}'' = \beta$. This observation, together with the fact that $\bar{u} > 0$, implies that the equatorial zone near the cloud tops has an absolute maximum of angular momentum. Theories of how this 'equatorial acceleration' is maintained against viscous dissipation are summarised elsewhere^{19,20}. A significant feature of the observations (Fig. 2) is that angular velocity increases poleward between 9° S and 9° N, although angular momentum decreases.

Texture and morphology

Jupiter at 100 km or better resolution exhibits a variety of textures, close-packed round puffy features, puffy features in linear arrays, short linear features, long curved or folded filaments, and amorphous regions in which only large-scale, regional variations appear. Texture is a regional characteristic that is distinct from brightness or colour, although certain ranges of brightness, colour and texture may tend to occur together. Using an Earth analogy, texture provides clues about processes that are occurring in the visible clouds. We aim here to develop a textural classification of jovian clouds, and to provide a global map based largely on texture.

Plate 8, p. 790 shows a montage of images showing examples of different textural types, together with our cartoon representation of each type. Type *a* includes round puffy clouds with irregular organisation, suggestive of small-scale cumulus convection on Earth. Type *b* includes puffy clouds in linear arrays, suggesting convection with lateral shear in the horizontal wind. Type *c* shows a linear pattern suggesting strong shear and/or waves with small irregular variations along the wave crests. Type *d* is an isolated bright region, usually with puffy clouds visible on the periphery¹; although this description includes a reference to brightness, these features, which appear suddenly and spread rapidly, are, nevertheless, a morphologically distinct type on Jupiter. Type *e* includes filaments—long, coherent, parallel threads suggesting laminar flow. Type *f* is a filamentary spiral, opening outwards either in a clockwise or counterclockwise direction. Type *g* is an example of folded filaments. Type *h* is a V-shaped pattern that forms where linear features with opposite tilts intersect. Type *i* is a region of low texture. Our cartoon representation of this type is simply a low density of lines, and is left blank.

Plate 9, p. 790 shows one face of Jupiter on 08.45 UT, 27 February 1979. This mosaic was made by projecting nine narrow angle frames in the orange filter onto a cylindrical grid. The resolution of the original frames is 120 km, although the resolution on the printed page is several times poorer. Plate 10a, p. 791 shows a morphological map of the same mosaic. The map was made from the nine orange frames and also from frames in other filters taken several days later at better than 100-km resolution. In the context of Plate 10a, the textural classification of any feature seems to be independent of the filter used in the images. At 150-km resolution, the texture of broad regions did not change during the several days before closest approach. We assume that the texture at higher resolution also did not change during this time interval. Studies of the highest-resolution (~ 8 km) frames suggest that the size of Jupiter's convective elements is at least 100 km, and therefore is adequately described in the following analysis.

Texture, like velocity, colour and albedo, is arrayed in a banded pattern around the planet. Bands at adjacent latitudes typically have oppositely tilted linear features, that is, features running north-west/south-east alternating with those running north-east/south-west. The intersections of these oppositely tilted features at adjacent latitudes form the chevron patterns (type *h*) that were mentioned in the first Voyager report¹.

Streamlines, defined by cloud motions, and streaklines, defined by the linear patterns, usually do not coincide. The latter are often tilted with respect to lines of constant latitude. The

former are rarely tilted; that is, the motion is largely zonal. From the pattern of zonal velocity versus latitude (Fig. 1) it is clear that the extremes of zonal velocity coincide with the vertexes of the chevrons. The vertexes ($>$ or $<$) point in the flow direction (eastwards or westwards, respectively). This chevron pattern occurs at all latitudes, although it is more common at mid-latitudes. The correlation with velocity structure implies that texture on Jupiter is a strong function of lateral shear in the horizontal wind. (But note that we have almost no information about vertical shear.) The linear patterns could represent either transient features that are being pulled apart by the shear between the jets, or waves that are generated near the jet maxima and propagate in a V-shaped pattern behind the jets.

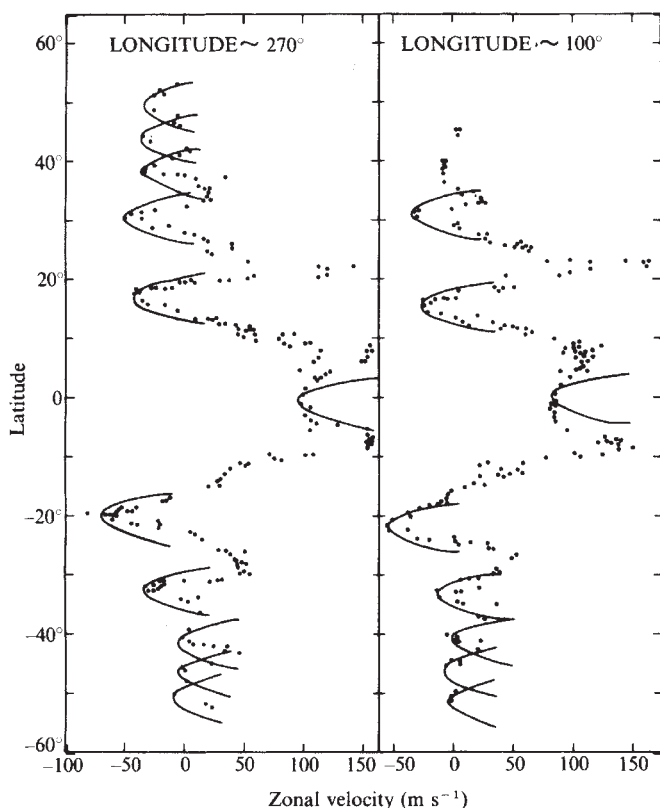


Fig. 2 Curvature $\bar{u}''$ of the zonal velocity profile compared to β . The measured points are the same as in Fig. 1. The parabolic curves are defined by $\bar{u}'' = \beta$, and have been centred on the westward jet maxima. At the equator one observes $\bar{u}'' < \beta$, which implies that angular momentum per unit mass decreases monotonically from equator to pole. At the westward jet maxima one observes $\bar{u}'' \geq 2\beta$, which implies that the absolute vorticity gradient with respect to latitude changes sign.

The Great Red Spot (GRS) and white oval immediately to the south also exhibit chevron patterns in the anticyclonic (counterclockwise in the southern hemisphere, clockwise in the northern hemisphere) flow around their peripheries. The inner parts of these chevrons give the GRS and similar spots an outward spiral appearance (for counterclockwise flow) at their centres and an inward spiral appearance outside. However, the direction of the spiral does not necessarily indicate convergence or divergence.

Rather, measured velocities around the GRS indicate predominantly circumferential motion. In fact, the angular velocity of rotation about the GRS centre seems to be greatest near the periphery². Evidence that convergence or divergence is small also comes from the long lifetime (~ 60 d) of a prominent white feature that circled in the GRS during the Voyager approach phase¹. In several cases, however, small spots were seen to enter the GRS peripheral flow from outside. No spots originating inside the GRS were seen to leave. This evidence seems to favour convergence, at cloud-top altitudes.

The chevron pattern also describes the direction of spirals at 34°N , 41°N and 41°S . If the material outside rotates more slowly than that inside, the former gets left behind by the rotation of the latter. This possibility is consistent with the observed direction of the spirals for these anticyclonic spots. Again, there is evidence that the inward spirals (for anticyclonic flow) are not streamlines and therefore should not be regarded as evidence of horizontal convergence.

The folded filament pattern (type g) is typical of the most rapidly varying, disorganised, turbulent regions. The lifetime of these features is about one jovian rotation. Oblong patches with this texture occur at 40°S separating the dark-haloed spots from each other. The oblongs typically lie somewhat equatorward of the spot centres. On a larger scale the same folded filament pattern occurs to the west of the white ovals and the GRS. These turbulent, folded filament regions always lie on the equatorward sides of the westward jets and have cyclonic vorticity. In contrast, the stable, well-defined spiral spots always lie on the poleward sides of the westward jets, and have anticyclonic vorticity (see Fig. 1 and ref. 2).

The equatorial region between 9°S and 9°N is generally characterised by low textural contrast. The plume heads are an exception as discussed earlier¹. In the plume tails, unlike mid-latitudes, the diffuse filaments seem to represent streamlines. The difference with mid-latitudes may reflect the greater deformations undergone by equatorial features during their lifetimes. That is, the plume heads act as concentrated sources of bright material which spreads to fill the much larger area of the plume tails. Any tilt with respect to the velocity vector is eventually eliminated by the shear itself.

Conclusions

The latitudinal profile of zonal velocity is more constant over long periods and more symmetrical about the equator than the albedo and colour profiles. Absolute vorticity does not increase monotonically with latitude; it decreases significantly at the latitudes of the westward jets. The texture at 100-km resolution follows the larger-scale banding. The most prominent textural feature is the chevron pattern centred on the zonal jets and on the high-speed circumferential flow around the GRS and other large anticyclonic ovals. The tilted linear elements of the chevron patterns are not streamlines, but reflect the effect of shear on the visible markings. The spiral markings also reflect the effects of shear, and do not necessarily imply convergent or divergent flow. Finally, there is a significant difference between anticyclonic and cyclonic regions. The latter are more compact and steady, forming well defined spots and vortices. The former are elongated in longitude and variable on time scales of one jovian rotation.

We thank the members of JPL's Image Processing Laboratory, particularly Glenn Garneau, Paul Jepsen and Gary Yagi for their assistance.

Received 16 July; accepted 27 July 1979.

- Smith, B. A. *et al.* *Science* **204**, 951-971 (1979).
- Mitchell, J. L. *et al.* *Nature* **280**, 776-778 (1979).
- Hunt, G. E. & Muller, P. *Nature* **280**, 778-780 (1979).
- Riddle, A. C. & Warwick, J. W. *Icarus* **27**, 457-459 (1976).
- Peek, B. M. *The Planet Jupiter* (Faber, London, 1958).
- Chapman, C. R. *J. Atmos. Sci.* **26**, 986-990 (1969).
- Smith, B. A. & Hunt, G. E. in *Jupiter* (ed. Gehrels, T.) 564-585 (University of Arizona Press, 1976).
- Beebe, R. F. *Nature* **280**, 771-772 (1979).
- Gehrels, T. in *Jupiter* (ed. Gehrels, T.) 531-563 (University of Arizona Press, 1976).

- Terrile, R. J. & Beebe, R. F. *Science* **204**, 948-951 (1979).
- Ingersoll, A. P. & Cuzzi, J. N. *J. Atmos. Sci.* **26**, 981-985 (1969).
- Ingersoll, A. P. *Icarus* **29**, 245-253 (1976).
- Kuo, H.-L. *J. Met.* **6**, 105-122 (1949).
- Rhines, P. B. *J. Fluid Mech.* **69**, 417-443 (1975).
- Williams, G. P. *J. Atmos. Sci.* **35**, 1399-1426 (1978).
- Williams, G. P. *J. Atmos. Sci.* **36**, 932-968 (1979).
- Gierasch, P. J., Ingersoll, A. P. & Pollard, D. *Icarus* (in the press).
- Pollard, D. & Ingersoll, A. P. *J. Atmos. Sci.* (submitted).
- Stone, P. H. in *Jupiter* (ed. Gehrels, T.) 586-618 (University of Arizona Press, 1976).
- Maxworthy, T. *Planet. Space Sci.* **23**, 1223-1233 (1975).

Jovian cloud structure and velocity fields

J. L. Mitchell*, R. J. Terrile*, B. A. Smith†, J.-P. Muller‡, A. P. Ingersoll§, G. E. Hunt‡, S. A. Collins* & R. F. Beebe||

* Jet Propulsion Laboratory, Pasadena, California 91103

† Department of Planetary Sciences, University of Arizona, Tucson, Arizona 85721

‡ University College London, London WC1, UK

§ Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125

|| Department of Astronomy, New Mexico State University, Las Cruces, New Mexico 88003

Regional views of the jovian cloud structures and associated flows demonstrate the asymmetry between the northern and southern hemispheres.

A GLOBAL view of jovian small-scale morphologies and zonal velocity profiles has been presented and discussed in an accompanying article¹. Here we give a regional comparison of the structures and velocity fields (meridional as well as zonal velocities) in the jovian atmosphere (scales >200 km) as observed by the Voyager 1 imaging system. In our descriptions we hardly mention their theoretical implications. All velocities are referenced relative to System III longitude. Cloud structures described are fairly independent of the visual wavelength at which the observation was made.

Although both hemispheres of Jupiter show similar patterns of diminishing and alternating eastward and westward jets as one progresses polewards, there is a pronounced asymmetry in the structural appearance of the two hemispheres. The southern hemisphere is dominated by bands of longitudinally alternating elliptical anticyclones and more elongated cyclones. The northern Hemisphere shows fewer longitude-dependent structures with the notable exceptions being the dark oblong features along the northern edge of the North Equatorial Belt (NEB) and a region of folded filaments in the NEB itself. The two hemispheres are separated by a structurally unique Equatorial Zone (EZ).

Northern hemisphere

Jupiter's northern hemisphere is characterised by a lack of large-scale (scales $\geq 10^3$ km) longitudinally dependent structures similar to those observed in the southern hemisphere. As seen by Voyager 1 the smaller northern features are, nevertheless, often similar in morphology and velocity structure to their larger, ubiquitous southern counterparts.

Cloud patterns in the NEB (latitudes from 9° to 18° N) vary considerably as a function of longitude. Diffuse equatorial clouds give way to a more clearly defined and organised structure at about 9° N. At this latitude a band of long filamentary clouds is present in a region of very strong shear between the equatorial current and a westward current at 18° N. Strong longitudinal differences in cloud structures also exist in the shear region. Between 40° and 100° W is a wedge-shape pattern of highly folded light and dark filaments rolling cyclonically (Plate 11, p. 792). This pattern is similar in morphology to the area west of the Great Red Spot (GRS) but it is not associated with any visible feature to the east or west. The northern pattern drifts independently of the similar southern pattern and has been observed many times before². Within this region small areas of bright puffy clouds about 500–2,000 km in diameter are observed to form and dissipate on time scales of a few jovian rotations. Note that this variable region extends over only 60° of

longitude. It is probably significant that such regions are observed only on the equatorward side of westward jets. Between 140° and 280° W at 17° N elongated dark brown features are present. These structures have generally sharp boundaries with adjacent lighter regions and show a pattern of linear striations inside. The vorticity of these features is, as expected from the profile of zonal velocity, cyclonic. Their association with 5 μ m warm regions as observed from Earth leads to the suggestion that these are clearer regions through which the darker lower level clouds may be seen (see Fig. 5 in ref. 3).

From 20° to 23° N, the clouds take on a tilted linear pattern, associated with the opposing currents at 18° and 23° N. Tilting north-east, these features form a chevron shape with north-west tilted rows of puffy clouds to the north. The apex of the chevron pattern lies in the strongest part of the eastward jet. These features from about 19° to 30° N are longitudinally periodic with a wavelength of about 500 km (Fig. 5 in ref. 3). So similar and rapidly varying are the linear features that they are difficult to identify from one rotation to the next (over a 10-h time interval).

At about 31° N a westward current exists within which are clouds of folded ribbon-like features. Farther north at about 35° N a distinct albedo change exists across a transition to an eastward current with darker clouds appearing to the north. Within this transitional area are small circular features about 3,000 km in diameter (34° N, 40° W; see Plate 11). These features are comprised of dark rings with white cores and exhibit anticyclonic rotation. Occasional spot interactions have been observed when one of these features overtakes another³. These features are not distributed uniformly in longitude. They are found spaced about 10°–15° apart in a longitude range of 350–260°W. They do not appear in the remaining 90° of longitude (Plate 11) but seem to be obscured by a higher level of diffuse clouds. Comparing Plates 11 and 12, p. 792 there is a similarity in the appearance of this band of spots and those at 41° S with regard to size and spacing of the individual spots. Interestingly, some of the spots at 34° N have been observed from Earth to be bright (hot) at 5 μ m, as have some of their southern counterparts. However, the spots at 34° N, which are anticyclones as are those at 41° S, do not have an alternating pattern of oblong cyclones associated with them as do the spots at 41° S.

The region from about 36° N to the northern limb (at about 80° N) is characterised by several white circular features rotating anticyclonically (40° N, 60° W; see Plate 11). In some cases faint spiral patterns can be observed in the interiors of the features. Surrounding these features are large areas of puffy unorganised clouds and other regions of folded ribbon-like clouds. A zonally elongated cyclonic flow stretching over tens of degrees of longitude (40° N, 30° W) is observed within these regions (Plate 10b and 11, pp. 791, 792).

Long currents are observed around these regions flowing eastwards along the southern edge, turning northwards, 'recirculating' in a westward flow, and then returning southwards. These recirculating cyclonic currents are similar to some of the ribbon-like flow patterns in the southern hemisphere, although

it is interesting that they are not associated with alternate anticyclonic spots as in the south. At least one small (2,000 km diameter) feature was observed at about 45° N, 55° W (*Plate 11*) to exhibit very rapid cyclonic rotation. Morphologically this feature appears similar to the recirculating currents except for its very compact size.

Equatorial zone

The uniqueness of the broad equatorial eastward jet and the characteristic structures which lie within is well known. At the time of the Voyager 1 encounter we found the EZ to lie between approximately 9° N and 9° S. Earth-based observations have shown that changes in the cloud structure of the region may occur rapidly, with the morphology of a specific feature changing radically within a single rotation of the planet.

At the time of the Voyager 1 encounter, the northern edge of the equatorial region was characterised by a train of 13 plumes encircling the planet³. The apparent regular spacing of these features strongly suggests wave propagation in the equatorial zone. In spite of this apparent longitudinal regularity of these features, there are some important differences in their individual morphologies. Only four of the 13 plume features were observed to have bright nuclei during the period covered by the Voyager 1 observations. It seems that there is no specific System III longitude at which changes were observed to occur within the plume heads. The high resolution images of the plume heads suggest that these albedo changes are associated with localised (scales ≤ 200 km) convective activity. Generally, the plume heads are uniformly bright over scales of 2,000 km, though smaller scale (100–200 km) bright features are occasionally observed on the edge of the plume head.

The plumes are observed to move eastwards with the strong equatorial jet with velocities ranging from roughly 100 to 150 m s^{-1} . We have, however, observed some longitudinal dependence in the zonal velocity profile across the EZ. Thus, not all plumes are observed to move eastwards at the same speed. A profile of mean zonal velocities (Fig. 1 in ref. 4) indicates that there is a relative maximum in the eastward jet at both the northern and southern boundaries of the region, while a relative minimum occurs at approximately the equator. The observed range in eastward velocities is from 90 to 150 m s^{-1} .

Note that the plume heads are found only along the northern boundary of the zone. A fundamental problem seems to be to explain this north/south asymmetry in the zone's morphology in light of the symmetrical zonal velocity profiles through the equatorial jet. It may be that activity within the South Equatorial Belt (SEB) influences the southern boundary of the EZ. Indeed, we have seen material pass from the SEB northwards into the equatorial jet. Small waves (scales 70–650 km) have been observed within the EZ and are discussed elsewhere⁴. Some of these waves are observed to persist as long as 14 jovian rotations. The appearance of these waves suggests some dependence on the camera filter used, though further processing is needed to confirm this suspicion. Groups of these waves are observed to move with eastward velocities typical of other features at the same latitude.

Southern hemisphere

In the southern hemisphere both the zonal velocity profiles and the observed cloud structures (compare *Plate 12*, p. 792 with Fig. 1 in ref. 1) are remarkably cyclic as one proceeds from equator to pole. As in the northern hemisphere the zonal velocity profile consists of alternating eastward and westward jets of decreasing magnitude and width as one proceeds poleward. In the south such an alternating jet profile was not previously observed from Earth because Earth-based observers must use the easily identified elliptical spots, which lie at latitudes not associated with the high-speed jets, to deduce drift rates. Indeed, the presence of the large elliptical spots in the southern hemisphere seems to inhibit the formation of small, high-contrast features such as those used by Earth-based

observers to deduce velocities in the northern hemisphere⁵. At three different latitudes (14° S, 31° S and 40° S) oblong cyclones with folded filament morphology are found to alternate longitudinally with the more clearly defined elliptical anticyclones. The cyclonic nature of these regions was not anticipated from Earth-based observations. *Plate 10b* depicts the observed flow near the longitude of the GRS. The observed flow in the region of the elliptical spots and the meridional profiles of zonal velocity at longitudes away from the spots are correlated on a variety of spatial scales. The three sets of elliptical spots (the GRS, white ovals and spots at 41° S) are found to lie at latitudes corresponding to the southside of strong westward jets (regions of anticyclonic shear), whereas the associated oblong cyclones lie at latitudes corresponding to the northside of these same jets (regions of cyclonic shear). Interestingly, the size of each of the elliptical spots seems to scale roughly as the speed of the associated jets between which it sits. The ellipticity of the elliptical spots remains fairly constant at each scale, although the spots seem to become rounder as one progresses polewards and as the jet speed diminishes. The same basic statements seem to apply to the oblong cyclones and the associated cyclonic shears which lie at the same latitudes.

The smaller scale morphologies within the elliptical spots are remarkably similar. One is struck by the general lack of organised structure in the larger spots' interiors, such as spiral cloud patterns typical of terrestrial tropical storms. The cloud texture in the interiors form chains of puffy clouds. In the case of the highest resolution mosaics of the GRS a limited number of what may be groups of convective cumulus cells are seen in several areas. Chevron morphologies around the outer edges of the GRS and white ovals suggest higher angular velocities around the spots' outer edges than in their interiors. Indeed, higher velocity jets around the spots' peripheries have been observed in time-series of Voyager images. Regions around the elliptical spots are for the most part darker than the spots themselves and are of a more filament-like texture (see Fig. 7 in ref. 3). In the case of the GRS and some of the spots at 41° S, the peripheries of the ellipse are particularly well defined by very dark collars of filamentary clouds located on the outer (cyclonic) side of the strong peripheral jets.

The morphologies associated with the oblong cyclones are quite different from those of the elliptical spots. When looking near the longitude of the GRS, one sees what resembles turbulent wakes west of the GRS, oval BC, and several of the spots at 41° S (*Plate 12*). Even smaller eddies along the northern and southern edges of the supposed wake can be seen; however, the notion of a wake is eliminated when the observer notes that the curl of these eddies does not change sign across the purported wake. As we have previously noted, velocity measurements immediately confirm that instead these are regions of cyclonic circulation.

At longitudes away from the GRS, the cloud texture within the SEB itself (latitudes between 9° and 23° S; see *Plate 12*) is generally rather diffuse with the strong eastward velocities of the equatorial jet slowly giving way to a much narrower westward jet along the south edge of the belt. However, on either side of the GRS the flows within the SEB are cyclonic. Why the cyclonic region in the SEB west of the GRS seems more turbulent than the analogous region to the east remains to be seen. It is along the southern edge of the SEB west of the GRS that the bright anticyclonic swirls often observed to interact with the GRS are formed.

The characteristic differences of the oblong cyclones between the three white ovals are interesting. We use here the classical nomenclature of Peek⁶. In *Plate 12* the ovals BC, DE and FA are located at longitudes of roughly 85°, 170° and 5° W, respectively. From oval BC to FA the cyclonic circulation is unmistakably outlined by a dark scalloped ribbon of clouds around a brighter, soft-textured interior. The analogous region between oval FA and DE is not nearly so pronounced, perhaps because of the vast 195° longitude spacing between these two ovals, while the cloud structures take on a folded filament appearance. The

appearance of this region and the vast gap between ovals FA and DE suggest a 'missing' white oval near 270° longitude where a peculiar isolated region of folded filaments is found (Plate 12). The western section of the region between ovals DE and BC is marked by a dark scalloped ribbon along the northern edge of a darker central axis. The eastern portion of the region is of a rather chaotic appearance with chevron-like filaments along an axis slightly south of the darker eastern section.

Note that the cyclonic regions associated with both the white ovals and the spots at 41° S are generally of two types: Type I, which resembles folded ribbon-like filaments (much like the region west of oval BC) and Type II, which shows lack of internal detail and a single dark outer loop of cloud filament (for example, the region between ovals BC and FA). Type I regions are seen at even higher latitudes in both hemispheres. A reasonable suggestion seems to be that the Type II cyclonic regions are just Type I regions veiled by a higher deck of white cirrus.

The band at about 41° S displays this same pattern of alternating elliptical spots just south of the oblong cyclones which lie equatorwards of the westward jet. In this case, however, the elliptical spots and associated oblong cyclones occur with approximately a wavenumber 12 periodicity. The elliptical spots themselves seem to be of two general types: the dark-haloed type, displaying a bright interior with little apparent structure and surrounded by a dark relatively broad peripheral collar, and the spiral type, displaying a spiral arm structure with arms originating near the spot's centre and extending for about 180° around the spot. A few of the dark-haloed variety have, in the past, been identified with 5- μ m bright (hot) regions from ground-based observations, leading some to suggest convective upwelling in the bright centre with downward flow in the dark halo⁷.

At more southerly latitudes there are suggestions of a continuation of the alternating jet pattern, though definitive measurements are yet to be made; however, one still observes something like the alternating pattern of oblong cyclones (mostly of Type I) and smaller elliptical spots.

Conclusions

The most striking characteristic of the planet as a whole is the asymmetry in cloud structures between the northern and southern hemispheres. Such a marked asymmetry is particularly odd in light of the similarity of the observed zonal velocity profiles in both hemispheres. We have observed that the northern hemisphere possesses some regions of strong lateral shear which do not give rise to any large-scale disturbances, whereas other regions of shear, such as the section of the NEB from 40° to 100° W, possess instabilities which have lasted over several years. In the southern hemisphere we have found that all elliptical spots seem to be anticyclonic (as indeed they are in the northern hemisphere) and are associated with longitudinally alternating regions of cyclonic flow lying slightly equatorward. The structures in the southern hemisphere are so regular as to suggest that the dynamics of the GRS, the larger white ovals, and even smaller spots observed further south are quite similar, though seen at differing scales. An obvious question posed by this observation is why should the GRS be red while other anticyclonic spots are generally white? We believe that the EZ is a region of localised intense convective activity. Again, we note a curious asymmetry in that convective regions all seem to lie along the northern edge of the EZ, though the zonal velocity profile is more nearly symmetric about the equator.

Historically, we recognise that the jovian cloud structures are capable of undergoing large-scale changes. We hope that the extended time scale offered by the Voyager 2 encounter in early July 1979 will provide some insight into the longer term climatic changes of Jupiter's atmosphere.

G.E.H. and J.P.M. are supported by SRC.

Received 16 July; accepted 31 July 1979.

1. Ingersoll, A. P. *et al. Nature* **280**, 773–775 (1979).
2. Terrile, R. J. & Beebe, R. F. *Science* **204**, 948–951 (1979).
3. Smith, B. A. *et al. Science* **204**, 951–972 (1979).
4. Hunt, G. E. & Muller, J.-P. *Nature* **280**, 778–780 (1979).
5. Beebe, R. F. & Youngblood, L. A. *Nature* **280**, 771–772 (1979).
6. Peek, B. M. *The Planet Jupiter* (Faber, London, 1958).
7. Terrile, R. J. & Westphal, J. A. *Icarus* **30**, 247 (1977).

Voyager observations of small-scale waves in the equatorial region of the jovian atmosphere

G. E. Hunt & J.-P. Muller

Laboratory for Planetary Atmospheres, Department of Physics and Astronomy, University College London, London WC1, UK

Wave-like disturbances in the equatorial region of the jovian atmosphere are observed to fall into two categories.

THE Voyager 1 imaging system has provided the first glimpse of a wide range of spatial scales of motion, over a period of four months, of the jovian atmosphere¹. These images have helped to characterise better the general and small-scale features of the atmospheric circulation. Historically, bright spots have been observed in the equatorial zone (EZ) for over 60 yr in Earth-based observations^{2,3}. Voyager 1 images show these bright spots to be similar to the features observed by the Pioneer missions⁴. They resemble plume-like objects (on a scale of 10,000 km), whose trails are parallel to the shear⁵. Voyager 1 images show a train of 13 plume features moving differentially, in an eastwards direction, along the northern edge of the EZ, between 9 and 1.7°N, four of which have high albedo nuclei (defined as active) in all visible wavelength filters. Figure 1 shows one of these

brighter plumes at high resolution, through a violet filter, taken seven days before Voyager 1's closest approach to Jupiter. The two boxes in the Fig. 1 outline the regions where small-scale (~600 km) periodic cloud structures⁶ were observed. The upper box shows these periodic cloud structures, which we believe are waves, with their axes perpendicular to the plume trails at a distance of 10,000 km from the plume nucleus. The lower box shows these waves with their axes perpendicular to the latitude circle of 1°S. We have studied 27 images which display these waves, and due to different contrasts⁶ through the filters, they have varying degrees of visibility. We classify these waves into two categories: the plume-tail (PT) waves which appear in the upper box of Fig. 1 have a 'horizontal' extent of up to 3,000 km and the cross-equatorial (CE) waves which appear in the lower box of Fig. 1 have a 'horizontal' extent of up to 20,000 km. These PT waves have been observed with varying intensity, to be associated with the same plume for some 14 jovian rotation periods⁶. During this time, no PT waves

appeared with any other plume and this particular plume became inactive after the plume nucleus separated and dissipated in the north equatorial belt (NEB). The PT waves appeared clearest in short-wavelength filters and were more pronounced when the plume nucleus was expanding. The CE waves appear at several locations south of the plume nucleus and in Fig. 1 are approximately 8,000 km due south of the nucleus. Figure 2 shows a wide angle violet filter image where the shadow of Io's disk appears near the CE waves. However, the passing of this shadow is unlikely to have any relationship to these waves as

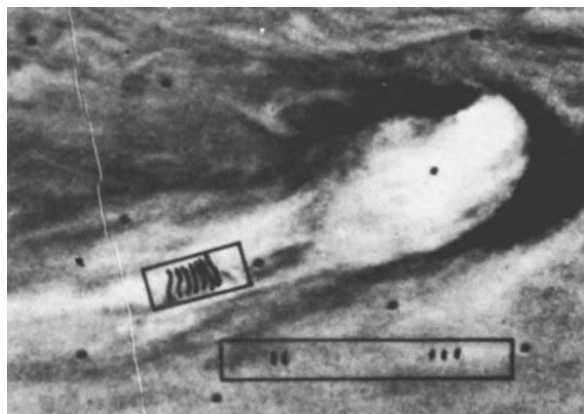


Fig. 1 A violet filter narrow angle of the area centred at 2.2°N and 89.98°W (System III) with a subspacecraft resolution of 136 km per television line pair. A train of waves is seen in the tail of the plume with a spacing of ~600 km. The cross-equatorial waves are only barely visible and their extent has been taken from the nearest orange filter frame. In all figures north is at the top.

the waves are visible when the shadow is not present. The CE waves have been observed in several forms for some 13 rotation periods in the same region of the EZ, though moving differentially to the plume drift rate.

The PT waves appear as alternate bright and dark lines with a spacing lying between 300 and 650 km (ref. 6) and appear to move at drift rates between 100 and 120 ms⁻¹, comparable to other trackable features in their vicinity including the plume nucleus (see Fig. 3). The CE waves appear in several forms including dark, isolated arcs and dark lines superimposed on a smaller-scale bright-and-dark pattern (see Fig. 4 which is a narrow angle enlargement of the area outlined in Fig. 2). The spacing between these lines varies between 70 and 650 km (ref. 6) and they appear to drift at rates between 90 and 110 ms⁻¹ which is comparable to the few other trackable features in this area (see Fig. 1^{4,7} and, for comparison, Fig. 3).

The PT waves appear to be the clearest when the plume nucleus is brightest. The variation in brightness of the plume nucleus has been interpreted to be consistent with terrestrial-

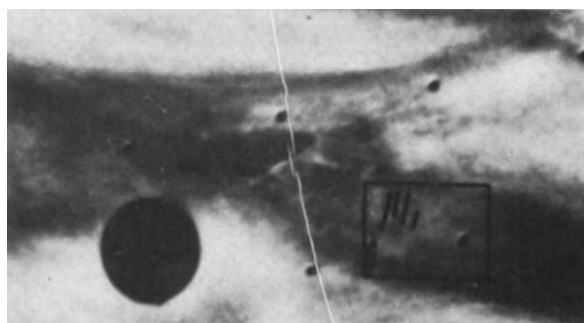


Fig. 2 A violet filter wide angle image of the same area as Fig. 1 taken some 13 rotation periods later, centred at 2.27°S and 166.60°W (System III) with a subspacecraft resolution of 139 km per television line pair. The cross-equatorial waves are visible inside the box with a spacing of ~600 km. The dark circular feature is the shadow of Io's disk.

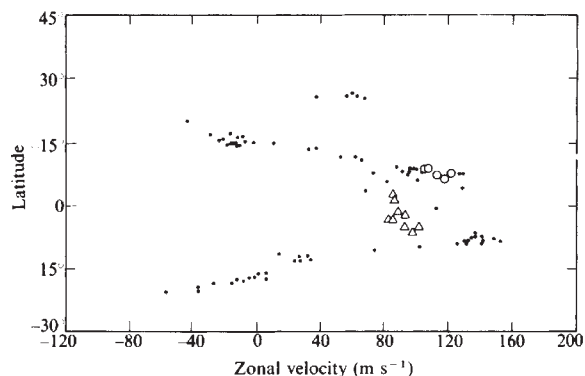


Fig. 3 The zonal velocity profile for the region occupied by both types of waves has been derived by tracking features from identical filter pairs with the image shown in Fig. 1 being the first of the image pair. The positions of the two types of waves have been shown. Zonal velocity has been derived with respect to System III longitude. ○, Plume tail waves; △, cross-equatorial waves.

type mesoscale convection¹. However, this phenomenon has also been observed in the turbulent regions of the NEB and south equatorial belt (SEB)⁶ though there is no evidence of any waves on this scale in these regions. The PT waves appear to be consistent with the interpretation that they are gravity waves initiated approximately 10,000 km upstream by the rapid, fluctuating convection in the plume nucleus. Disturbances of this kind have been observed in the Earth's atmosphere following severe thunderstorms⁸. Gravity waves have also been found to be precursors of strong convective activity in the terrestrial atmosphere⁹. This suggests that the PT waves may be directly related to the renewed brightenings of the plume nucleus in the jovian atmosphere and they may be an example of nondispersive gravity waves. Furthermore, type 'a' texture⁵ has been observed in the vicinity of the plume tail, supporting the contention that small-scale convection is present in this region. The CE



Fig. 4 A green filter narrow angle image taken two minutes before Fig. 2 and showing an enlargement of the box outlined in Fig. 2. It is centred at 1.4°S and 163.85°W (System III) with a subspacecraft resolution of 20 km per television line pair. The bright waves have a spacing of ~70 km and the dark waves of ~300 km.

waves, however, do not appear to have an obvious source mechanism.

Wave-like disturbances in the equatorial region of the jovian atmosphere have been shown to fall into two distinct categories differentiated by their 'horizontal' extent, morphology, spacing, drift rate and association with other visible features. It is hoped that further evidence to understand their dynamical significance will be gained from Voyager 2 observations in progress.

This research is supported by the SRC.

Received 19 July; accepted 31 July 1979.

1. Smith, B. A. *et al. Science* **204**, 951-972 (1979).
2. Peek, B. M. *The Planet Jupiter* (Faber, London, 1958).
3. Smith, B. A. & Hunt, G. E. *Jupiter* (ed. T. Gehrels) 564-595 (University of Arizona Press, 1976).
4. Fimmel, R. O., Swindell, W., Burgess, E. *Pioneer Odyssey* (NASA SP-396, 1977).
5. Ingersoll, A. P. *et al. Nature* **280**, 773-775 (1979).
6. Mitchell, J. L. *et al. Nature* **280**, 776-778 (1979).
7. Chapman, C. R. *J. Atmos. Sci.* **26**, 986 (1969).
8. Erickson, C. O. & Whitney, L. F. *Mon. Weath. Rev.* **101**, 708 (1973).
9. Uccellini, L. W. *J. Atmos. Sci.* **103**, 497 (1975).

A lower limit on the top of Jupiter's haze layer

A. F. Cook II*, T. C. Duxbury† & G. E. Hunt‡

*Center for Astrophysics, Harvard College Observatory and Smithsonian Astrophysical Observatory, Cambridge, Massachusetts 02138

†Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91103

‡Laboratory for Planetary Atmospheres, Department of Physics and Astronomy, University College London WC1, UK

Remote sensing observations of the atmosphere at wavelengths from the UV to the IR are affected by the presence of haze layers above the visible clouds. Such layers are difficult to detect as they generally contain small particles ($\leq 1 \mu\text{m}$). An imaging observation of high-altitude haze is presented that extends through the jovian stratosphere into the mesosphere.

THE observation was taken by the wide-angle camera of Voyager 1 at 1979 March, 4 d 17 h 52 min 36 s UT, to search for aurorae, lightning, and fireballs (Fig. 10, ref. 1). A star was occulted by Jupiter during the initial slew of the scan platform to the pointing direction for the first of a pair of images used of that field. Two other stars appear with sufficient brightness to leave trails during the slew. These trails, along with the measured positions of all of the stars, offer the possibility of looking for atmospheric refraction of the occulted star. The detailed image pattern of the stars is found both for the images that drifted slowly during the first pointing of the double exposure and of the streak left during the slew. The fit to the occulted stars is perfect, indicating that the double horizontal refraction at the top of the layer of haze did not exceed 14 arc s. We shall show that this corresponds to an upper limit of 3.5 mbar for the atmospheric pressure at the top of the occulting haze layers if the scale height is 30 km. This upper limit is proportional to the 3/2th power of the scale height.

Thirteen stars, ranging in visual mag from 6.8 to 9.0, were used to predict the location of the 7th mag star which was occulted. A photogrammetric approach was used with raw image locations (sample and line number) of stars and reseaux as the observations. Reseaux image locations were used to estimate local TV geometric distortions (up to 3 pixels) for the 14 stars. Star image locations were used to estimate TV camera pointing, orientation, and focal length. Inertial positions of the stars were obtained from the AGK3 catalogue.

Star image locations were measured manually from computer listings of imaging data surrounding each star. Star images, rather than being point sources, had been smeared about 10 samples and 20 lines due to camera slewing and spacecraft limit-cycle motion during the 3-min and 12-s exposure. Image smearing limited the measurement accuracy to about 1 pixel (14.0 arc s).

Using all the parameters estimated from the reseaux and 13 star image locations provided a prediction for the image location of the occulted star. It was found that the observed image location of the star up to the point of occultation agreed with the predicted position within the measurement uncertainty (Table 1).

The difference between the observed and computed image locations (sample and line number) of 13 stars plus the occulted star is listed in Table 1. This difference for the occulted star is within the measurement accuracy (~ 0.8 pixels) of the 13 other stars.

Figure 1 shows the fit of an outline of the slew and drift pattern taken from the brightest other star and superimposed on the computer pixel listing for the occulted star. The outline of the limb is also displayed. Clearly even 1 pixel of refraction would have visibly distorted the slew. Furthermore, failure to recognise such a distortion and a corresponding forced fit would lead to an enlarged systematic error in the deduced position of the star, which did not occur. One pixel corresponds to 14 arc s at the scale of the wide-angle camera and we thus arrive at an upper limit for the double horizontal refraction at the top of the jovian haze layer, 14 arc s.

The double horizontal refraction, θ , can be related for an isothermal atmosphere to the refractivity, ν_1 , at the minimum height of the ray and the scale height, H (ref. 2) by

$$\theta = \nu_1(2\pi a/H)^{1/2} \quad (1)$$

where a is the radius of the planet.

The refractivity is proportional to the density and we let the constant of proportionality be denoted by κ . Application of the perfect gas law enables us to find the pressure, P , at the minimum height of the ray:

$$P = \left(\frac{H^3}{2\pi a}\right)^{1/2} \frac{g\mu m_A}{\kappa k} \theta \quad (2)$$

where g is the acceleration due to gravity, μ the mean molecular weight, m_A the atomic mass, and k the Boltzmann's constant. The definition of a scale height is also used:

$$H = \frac{kT}{\mu m_A} \quad (3)$$

Table 1 Predicted star location errors

AGK3 no.	Visual mag	Spectral class	Predicted sample	Location error line
105363	8.9	K2	-0.6	0.0
205764	8.2	G5	0.1	0.1
105365	8.6	F0	0.5	-0.1
105370	8.4	K2	1.6	-1.7
205785	8.9	F8	-0.6	-0.1
205789	8.4	K0	-0.6	-0.0
205792	8.7	F0	-0.1	0.2
205796	9.0	G0	-1.6	1.4
205795	7.9	K0	-0.0	0.5
105378	9.0	K2	0.2	1.0
205814	8.8	A2	0.2	-0.0
205815	8.7	G5	0.4	-0.5
205818	6.8	G5	0.3	0.1
205766*	7.0	A0	0.5	-1.0

* Occulted star

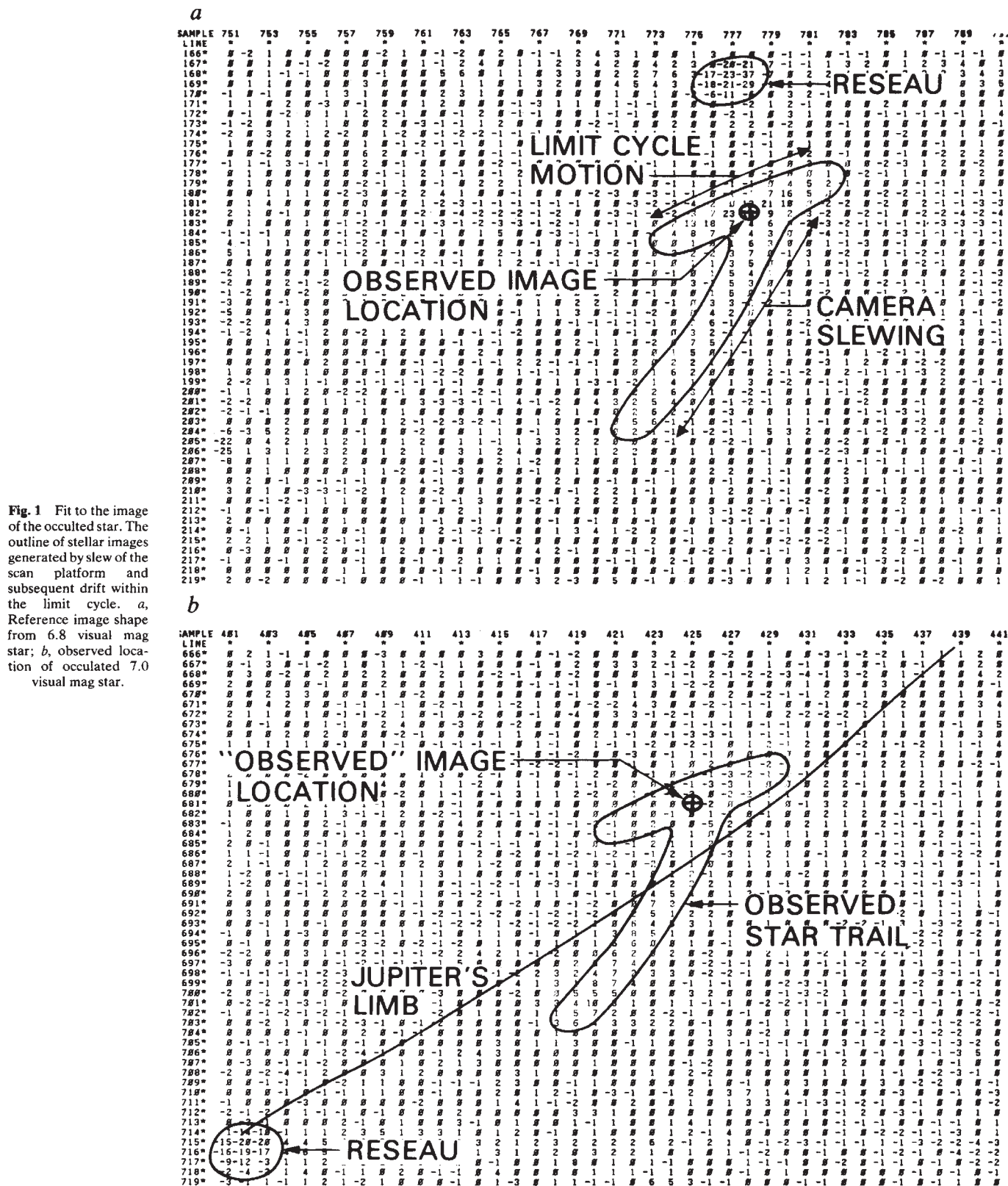


Fig. 1 Fit to the image of the occulted star. The outline of stellar images generated by slew of the scan platform and subsequent drift within the limit cycle. *a*, Reference image shape from 6.8 visual mag star; *b*, observed location of occulted 7.0 visual mag star.

where T is the absolute temperature. At a scale height of 30 km we find an upper limit on the pressure of 3.5 mbar. This very low value applies at jovian longitude $+35^\circ$ and latitude $+80^\circ$ and indicates that the haze layer extends far above the visible clouds of Jupiter into the mesosphere. It is therefore important to determine the structure and opacity of the haze in order to accurately interpret spectroscopic observations of Jupiter at wavelengths from the UV to the infrared.

These results were made possible by all those who have worked in the Voyager Project, many of whom are individually acknowledged in ref. 1. We thank especially C. C. Avis, G. W.

Garneau, P. L. Jepsen, J. J. Lorre, J. A. Mosher, G. M. Yagi, J. T. Harwood, L. Pieri, and C. Reslock of JPL, and also S. A. Collins as experiment representative at JPL. This paper reports one phase of research supported by NASA at JPL under contract NAS7-100 and supported by JPL at the Smithsonian Astrophysical Observatory under that NASA contract. G.E.H. was supported by the SRC.

Received 19 July; accepted 31 July 1979.

- Smith, B. A. *et al.* *Science* **204**, 951-972 (1979).
- Hunten, D. M. & Veeverka, J. in *Jupiter* (eds Gehrels T. & Matthews, M. S.) 247 (University of Arizona Press, Tucson, 1977).

Plate 1

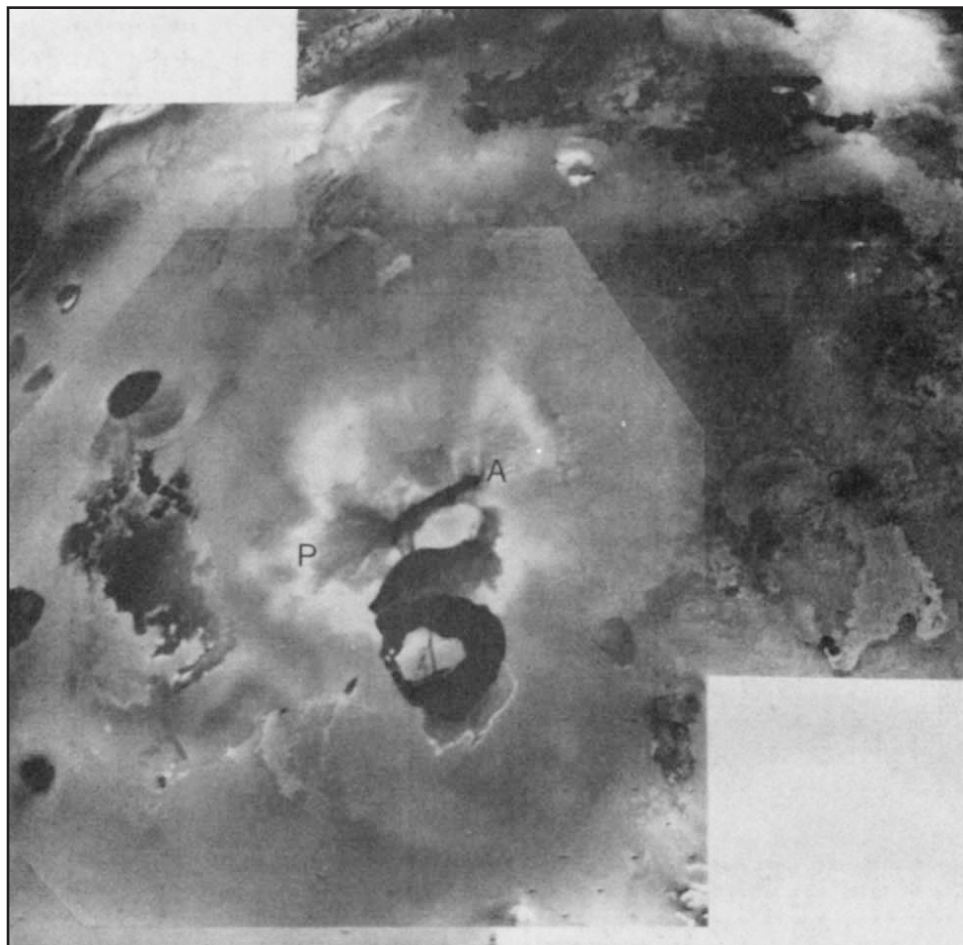


Plate 1 High resolution (2.7 km per television line pair) picture of the eruption site of Plume 2. The plume is probably a fissure eruption occurring along the length of the black strip A. The principal activity seems to be concentrated at the west end of the black area where enormous jets of pyroclastics (P) are being ejected. (FDS 16389.38).

Plate 2 *a*, Flow patterns in the region of 15°S, 325°W. At right centre several long, narrow dark flows originate at dark caldera. Several of the flows apparently have bright margins. In the lower left a dark-floored caldera is near the centre of some radial bright markings that could also be flows. Diffuse bright and dark markings are associated with most other calderas. The entire scene is ~900 km across (Frame 0043J1 + 000). *b*, Bright flow in middle right originating in small dark caldera, about 250 km long. It apparently has inundated at least one other caldera. Also shown are several discrete dark calderas, about 100 km in diameter. Near the centre of the frame is a mottled area surrounded by at least 10 broadly lobate flow-like markings of low albedo described in the text (Frame 0105J1 + 000).

Plate 3 The layered terrains of the south polar region. The uppermost layer in the stack at the upper left corner of the picture shows a complex erosion pattern. The layer immediately below is even more eroded and in the upper centre of the picture breaks up into both flat and smooth topped remnants. This area also contains long sinuous erosion troughs. In the lower centre of the picture a complex array of caldera-related bulbous flows overlies a stack of flat-topped layers that show the same type of erosional pattern observed in the south polar region. Bright streaks occur along the lowermost erosional scarp. Emanating from beneath another eroded scarp is a 100 km dark, digitate flow, the distal end of which is surrounded by a bright aureole. The scarp-related bright plumes are believed to be SO₂, the exhalation of which is driving the scarp retreat process. The long dark digitate flow may consist of sulphur released in liquid from the base of the stack.

Plate 4 *a*, Picture of irregular mountainous terrain that may be cut by faults and marked by craters, probably of internal origin. The mountains are surrounded by layered deposits that may have been deposited as ash flow tuffs. The arcuate scarps may define a caldera ringed by old mountains and layered plains deposits. *b*, Picture of volcanic vent surrounded by dark ejecta that resembles lunar dark halo craters. This crater lies on the flank of the source vent of Plume 1. Mountainous terrain that may be broken by faults is surrounded by layered deposits. This area, covered by young volcanoclastic deposits, shows simple geologic patterns in contrast to the complex patterns in *Plate 2*.

Plate 5 *a*, Voyager 1 image of Io (left) showing volcanic vents, flows and domes, and geologic map (right) showing interrelationships among terrain units. The image was acquired from a range of 133,000 km and covers an area approximately 1,000 km on a side. The geologic symbols are the same as those in Fig. 1 of Masursky *et al.* (this issue). md, Dark mantle material interpreted as volcanic debris cast from volcanic centres; c, volcanic cones; l, bright (high-albedo) material perhaps fumerolic deposits of sulphurous compounds. Lava flows or other fluid materials occurring in association with pit craters: d, dark; pcl, bright; pcm, mottled; and pcu, mixed-albedo; pck, knobby material associated with dark pit craters. Flow materials associated with volcanic shield-like forms: sd, dark; sl, light; sm, mottled; and su, undifferentiated subunits; sk/sh, knobby and hummocky materials on or near shields. *b*, Comparison of the eight eruptive plumes photographed by Voyager 1. The bar is 200 km long. The top of Plume 6 is visible to the right of Plume 5 and the top of Plume 5 is visible to the left of Plume 6.

Plate 6 *a*, Voyager 1 photograph (FDS 16377.50) showing the location of the eruption sites associated with Plumes 1, 4, and 7, and Plume 3 seen in profile at the limb. The width of Plume 3 is about 250 km. *b*, High resolution (1.5 km per television line pair) oblique picture of the eruption site of Plume 1. The dark sky has been specially processed to show the filamentary structure at the top of the plume. Measurements of this image indicate the plume filaments are about 280 km above its source, consistent with measurements on other images. The three main eruption sites responsible for the dark radial streaks are near the centre of the picture. A caldera is apparently related to the largest eruptive centre. Smaller diffuse dark areas are eruption sites which may have been active when the picture was taken. All eruption sites are associated with a highly fractured uplift about 200 km wide. (FDS 16391.30).

Plate 7 *a*, The eruption sites of Plumes 3, 4, and 7 are shown on this Voyager 1 photograph (FDS 16375.28). Plume 6 is shown in profile on the limb. Its central fountain is about 60 km wide. Dark jets of pyroclastic material are being ejected on ballistic trajectories at the site of Plume 3. To the right of A is the region where changes have occurred during the 5.5 h interval since *b* was taken. Many of the dark spots and bands seen in *b* have almost completely disappeared in this picture. *b*, Full disk picture (FDS 16368.28) of Io showing the locations of five eruption sites associated with Plumes 3, 4, 5, 6, and 7 (arrowed numbers). Unnumbered arrows indicate the probable sites of recent eruptions. The appearance of the region to the right of A has changed significantly since this picture was taken. Many of the dark bands and spots have disappeared in *a* taken 5.5 h later. The outer diameter of the bright ring surrounding site 3 is about 400 km.

Plate 8 Textural types observed on Jupiter. Each image is approximately 7,000 × 7,000 km, and was taken through the violet filter. The inset shows the cartoon representation of each textural type, as used in *Plate 10a*. These types are discussed individually in the text.

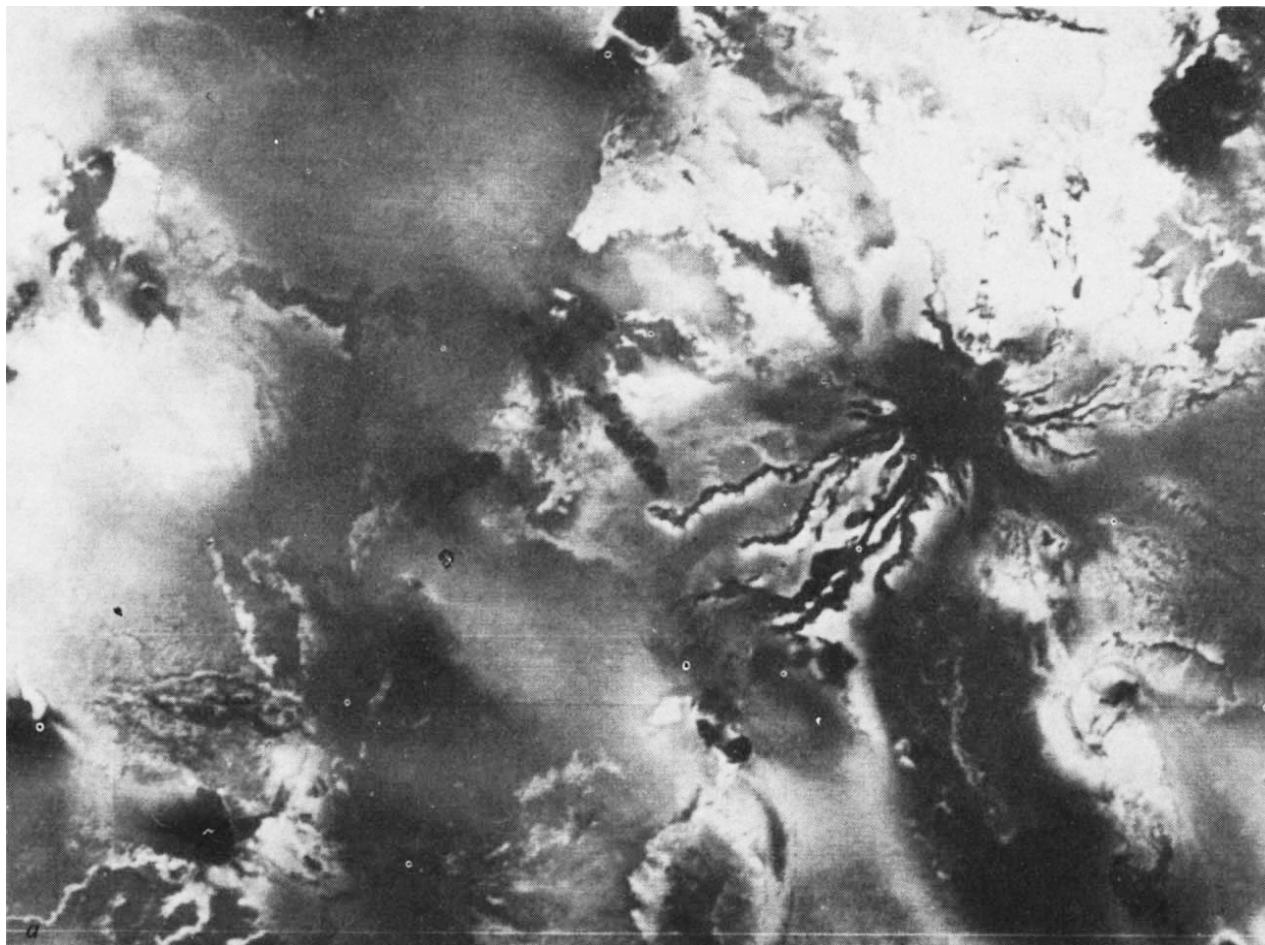


Plate 2

**Plate 3**

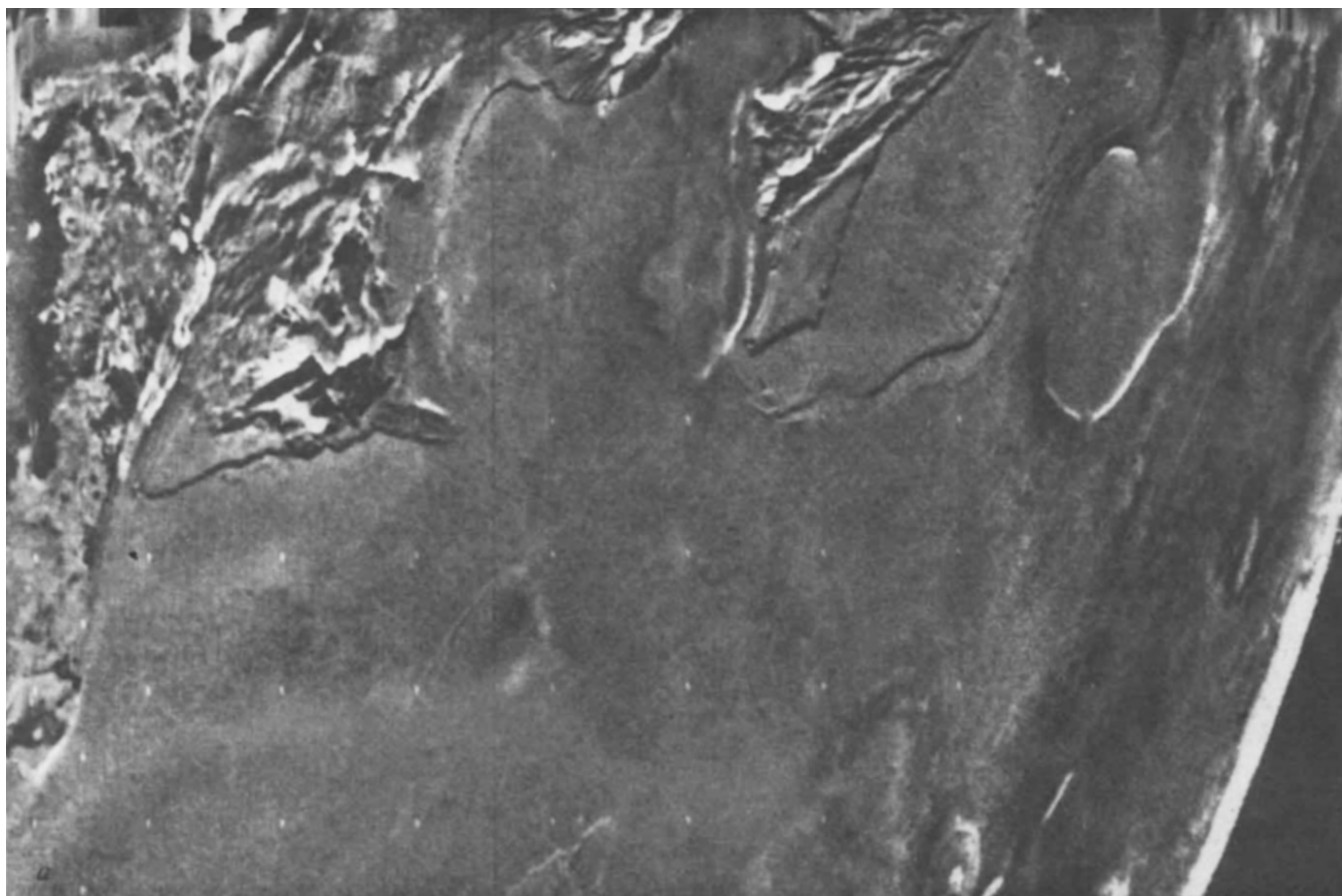


Plate 4

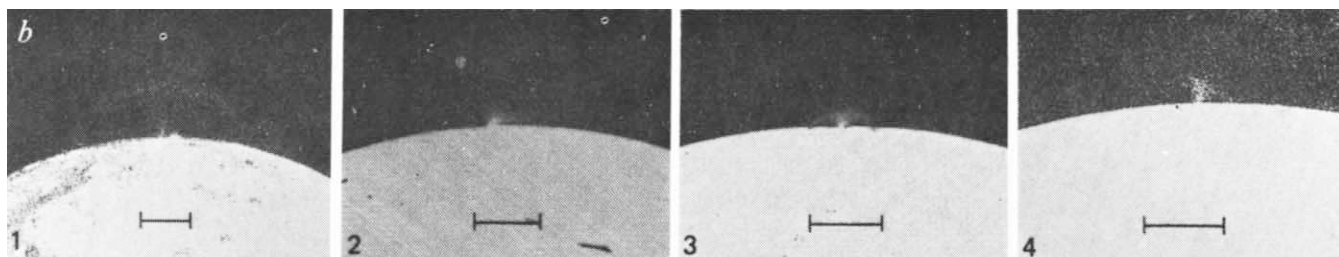


Plate 5

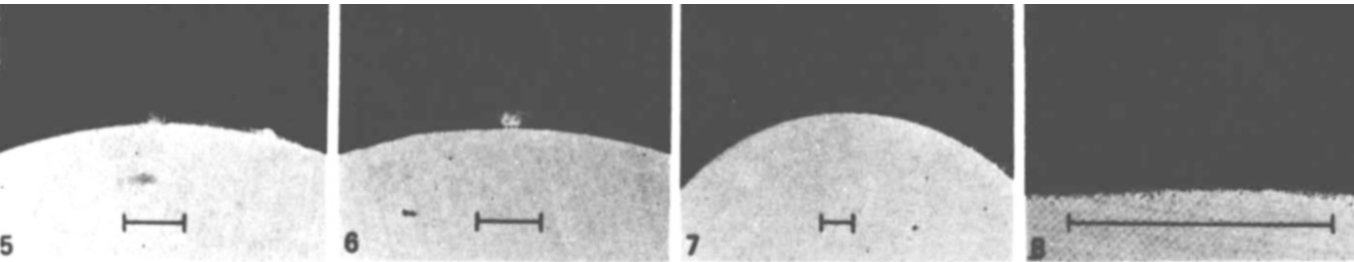
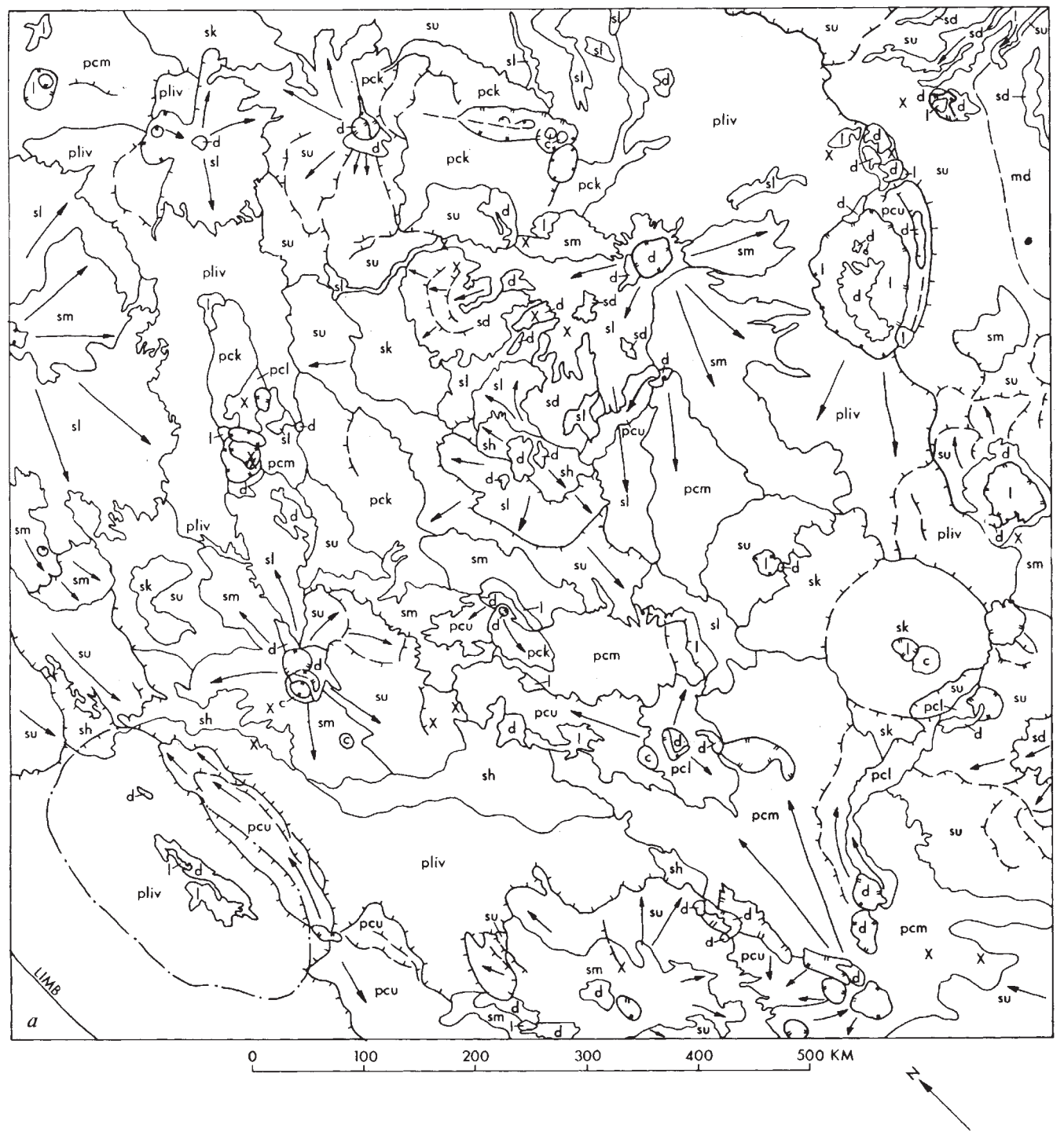
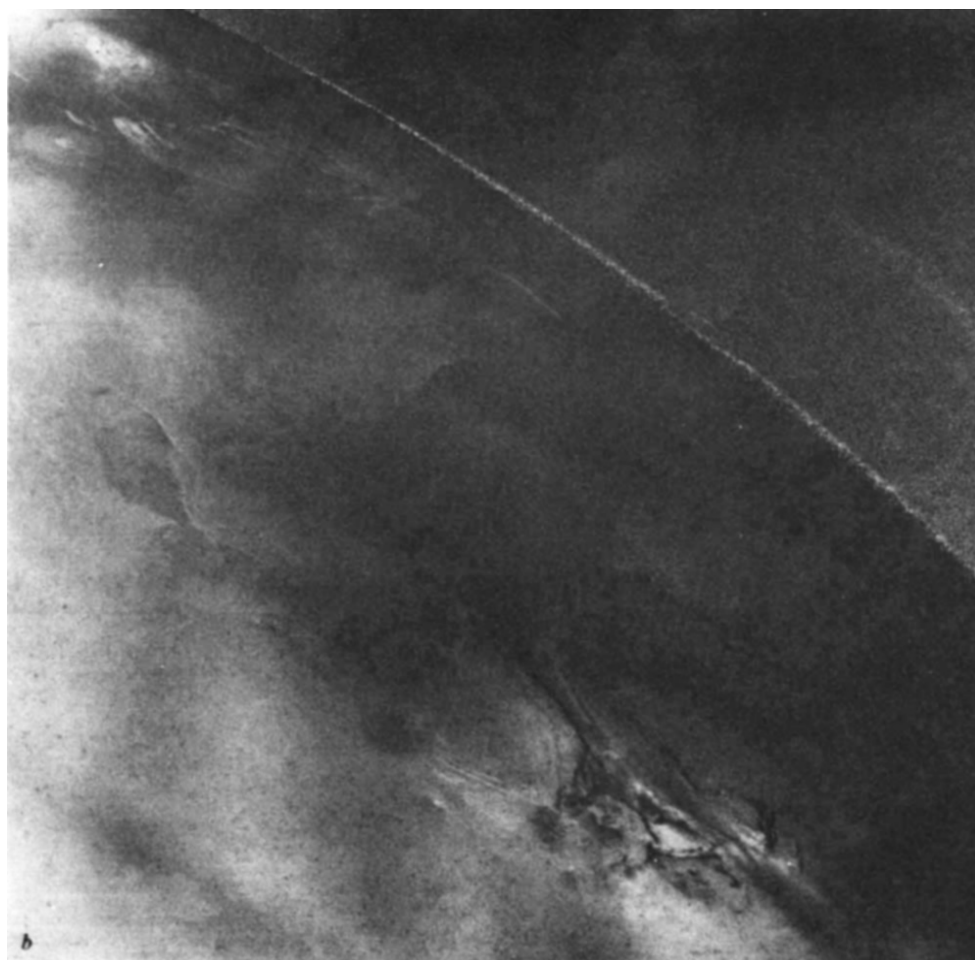
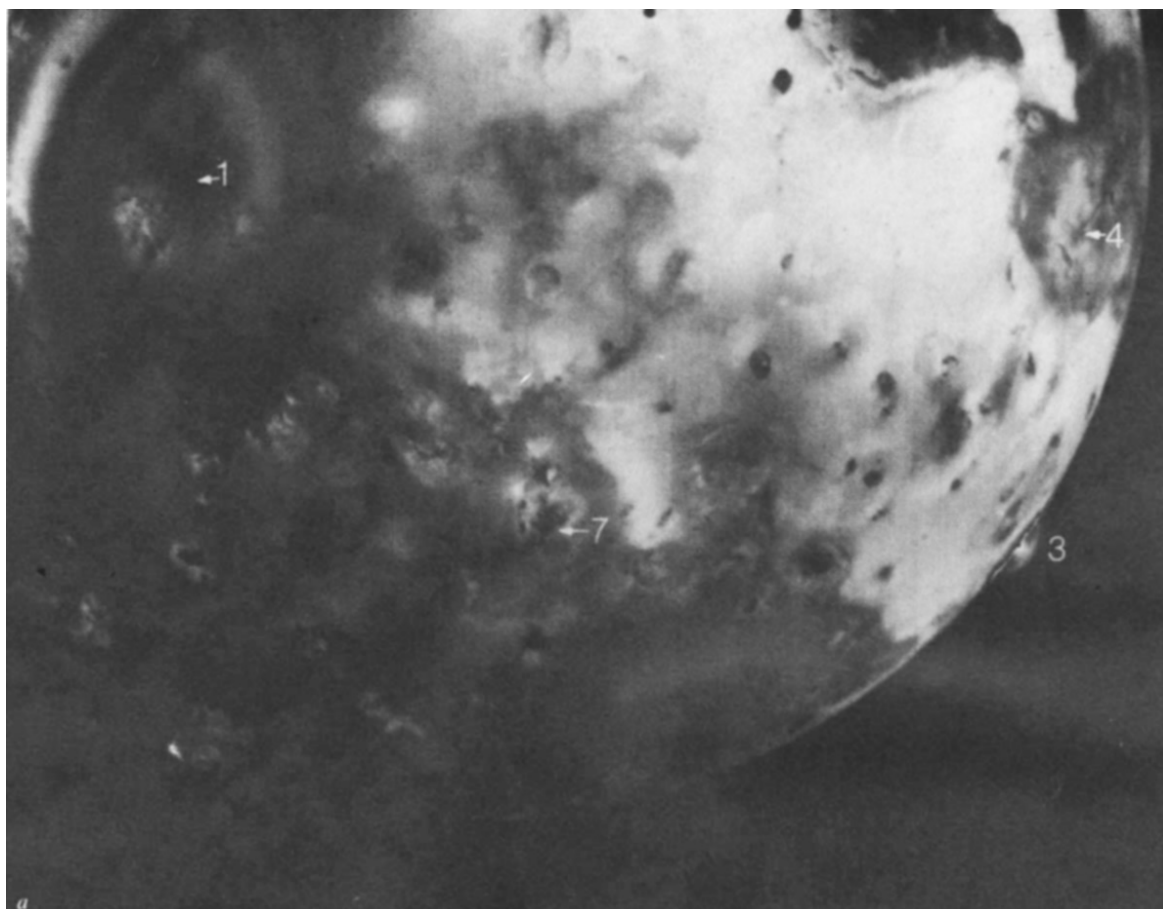


Plate 5

**Plate 6**

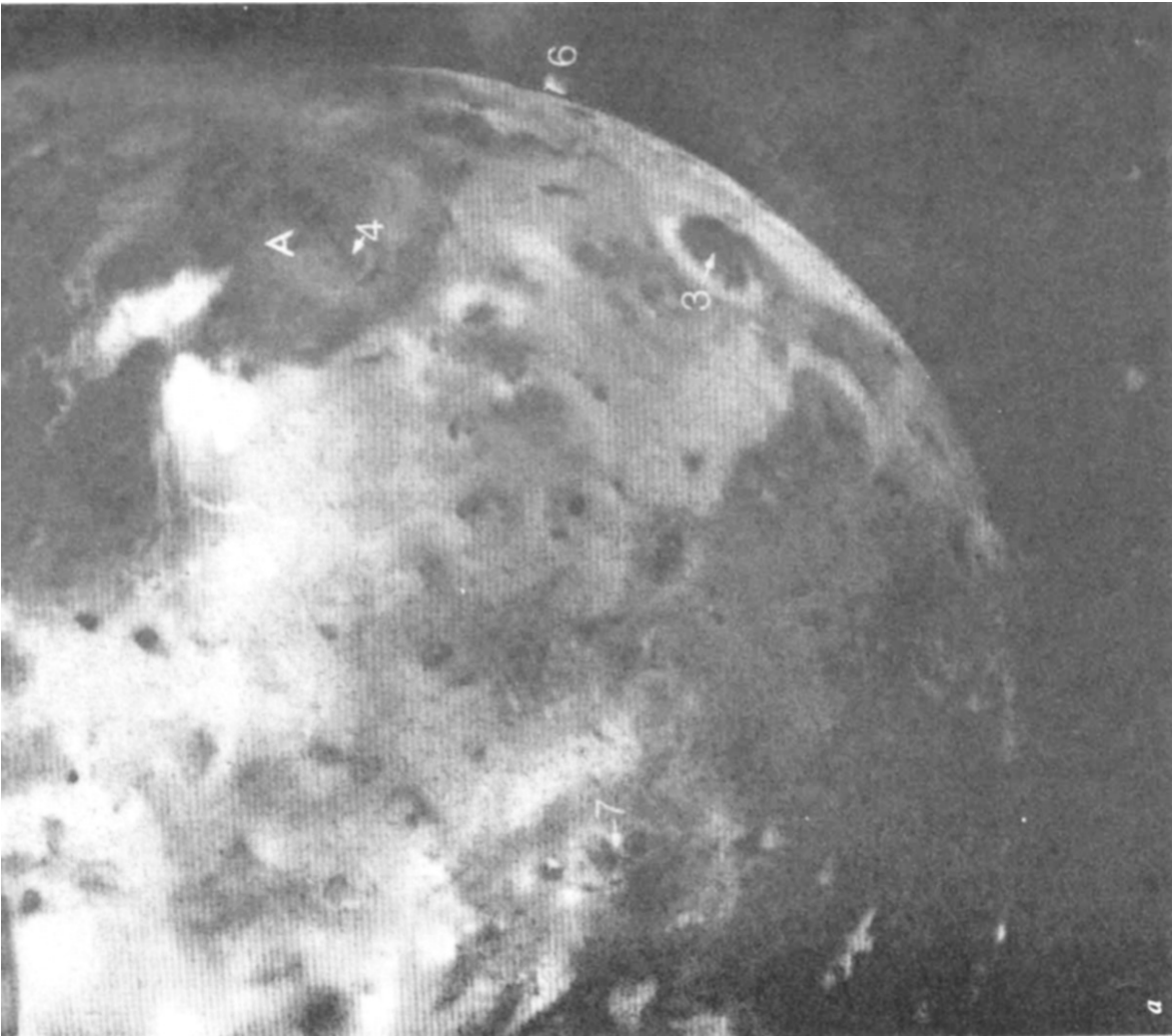
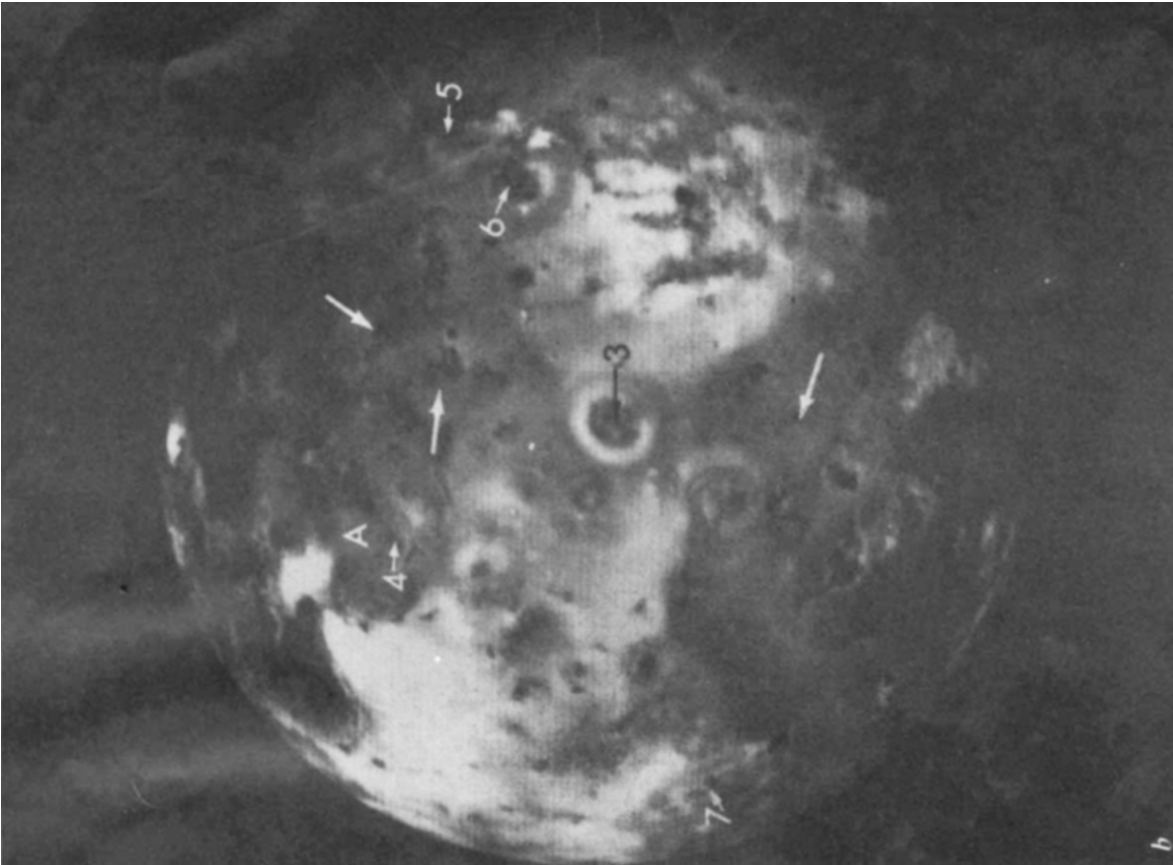


Plate 7

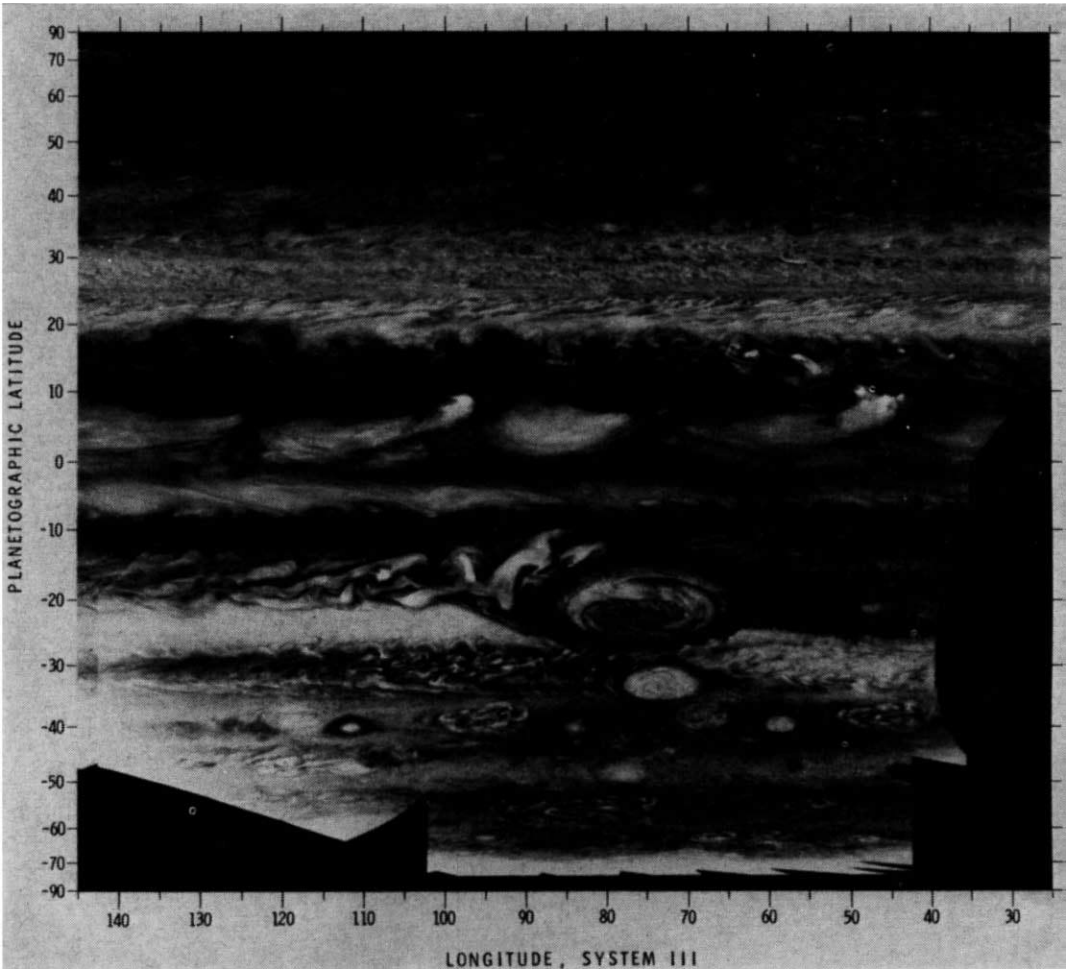


Plate 9

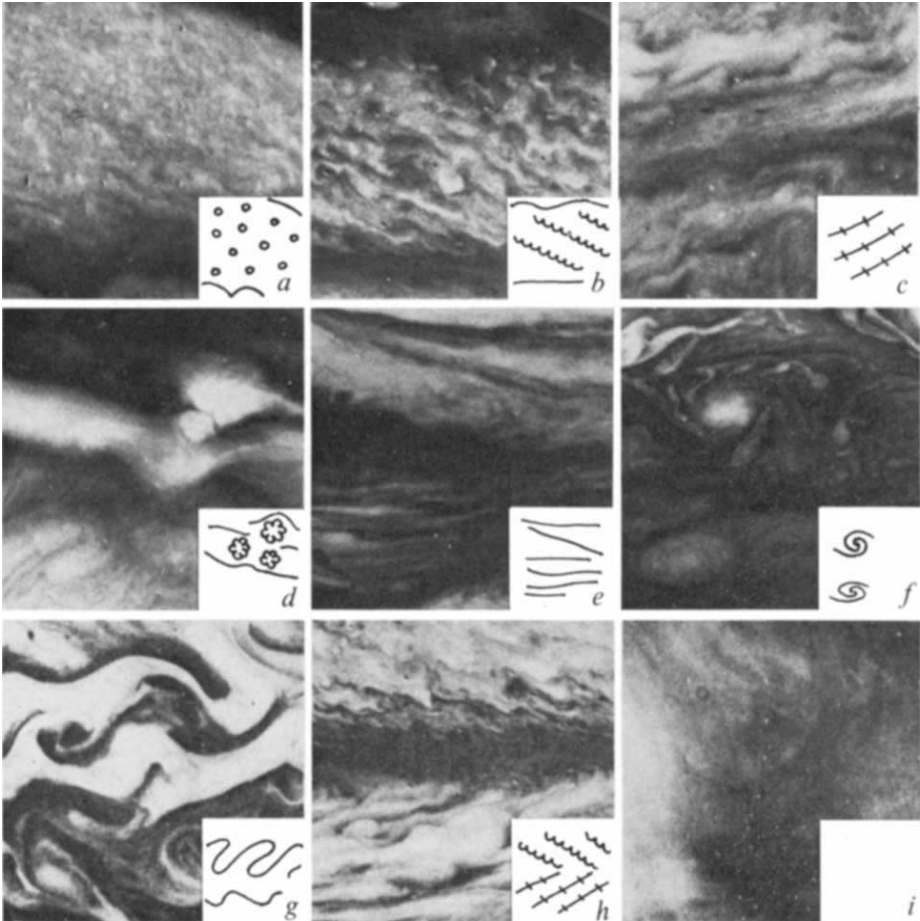


Plate 8

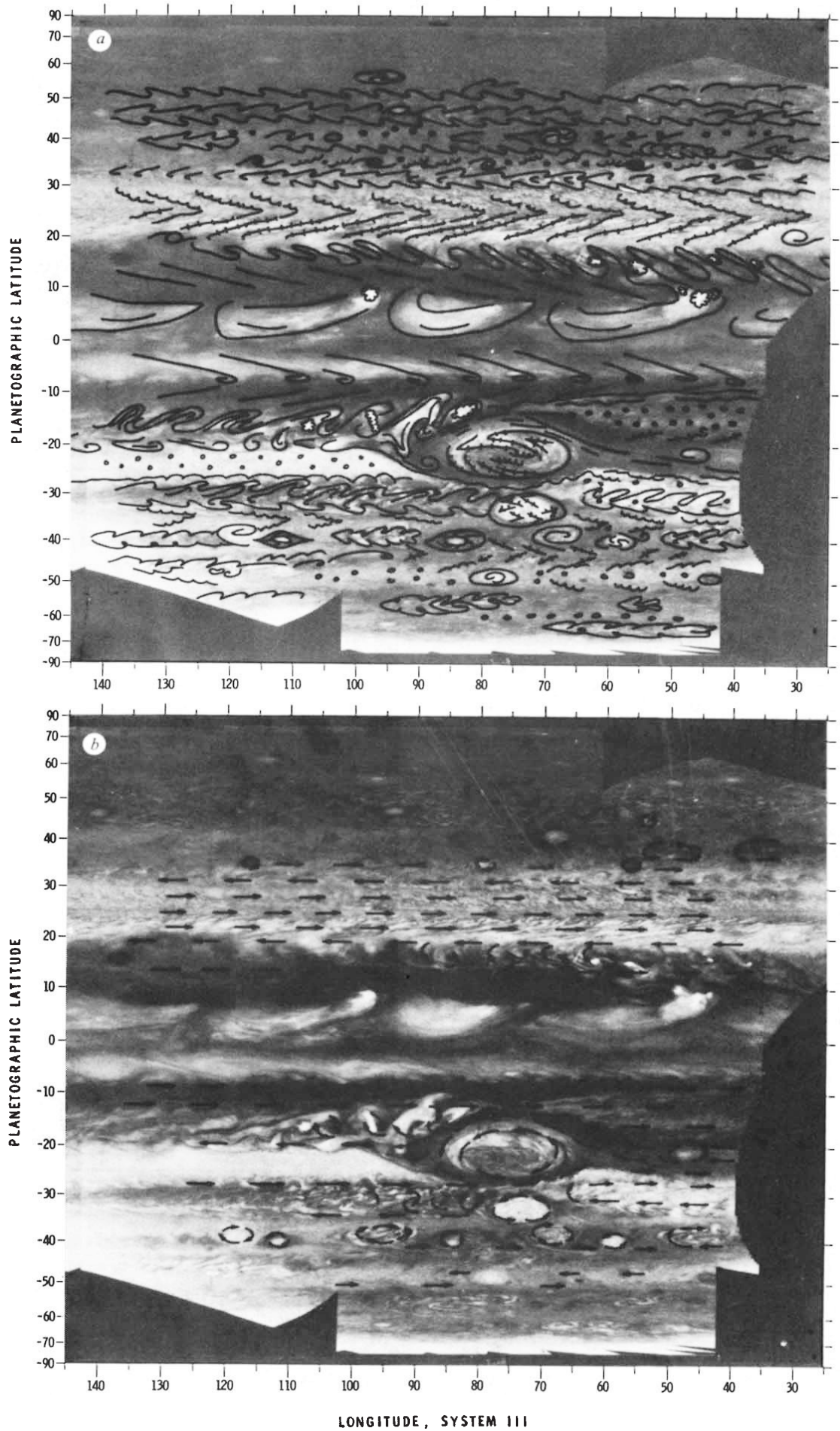


Plate 10

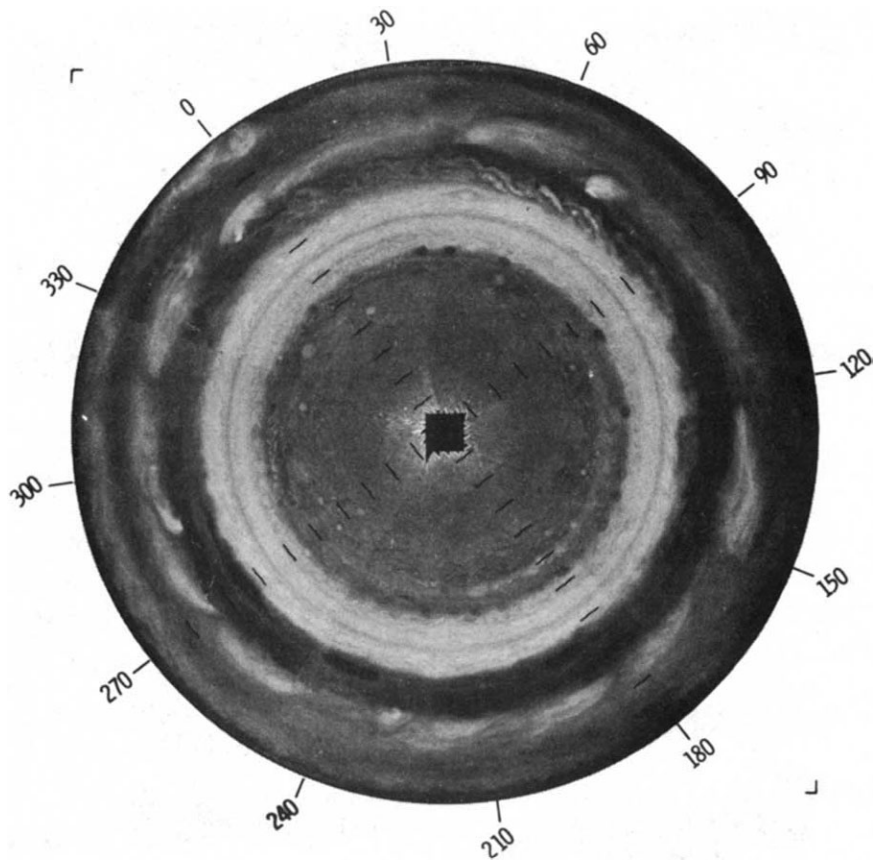


Plate 11

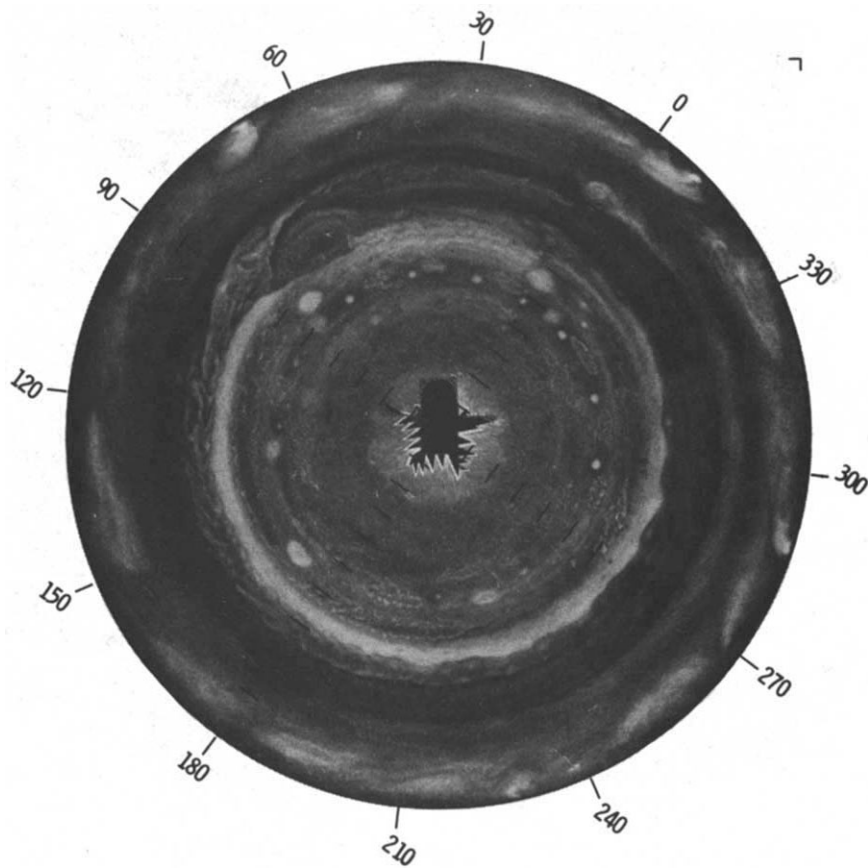


Plate 12

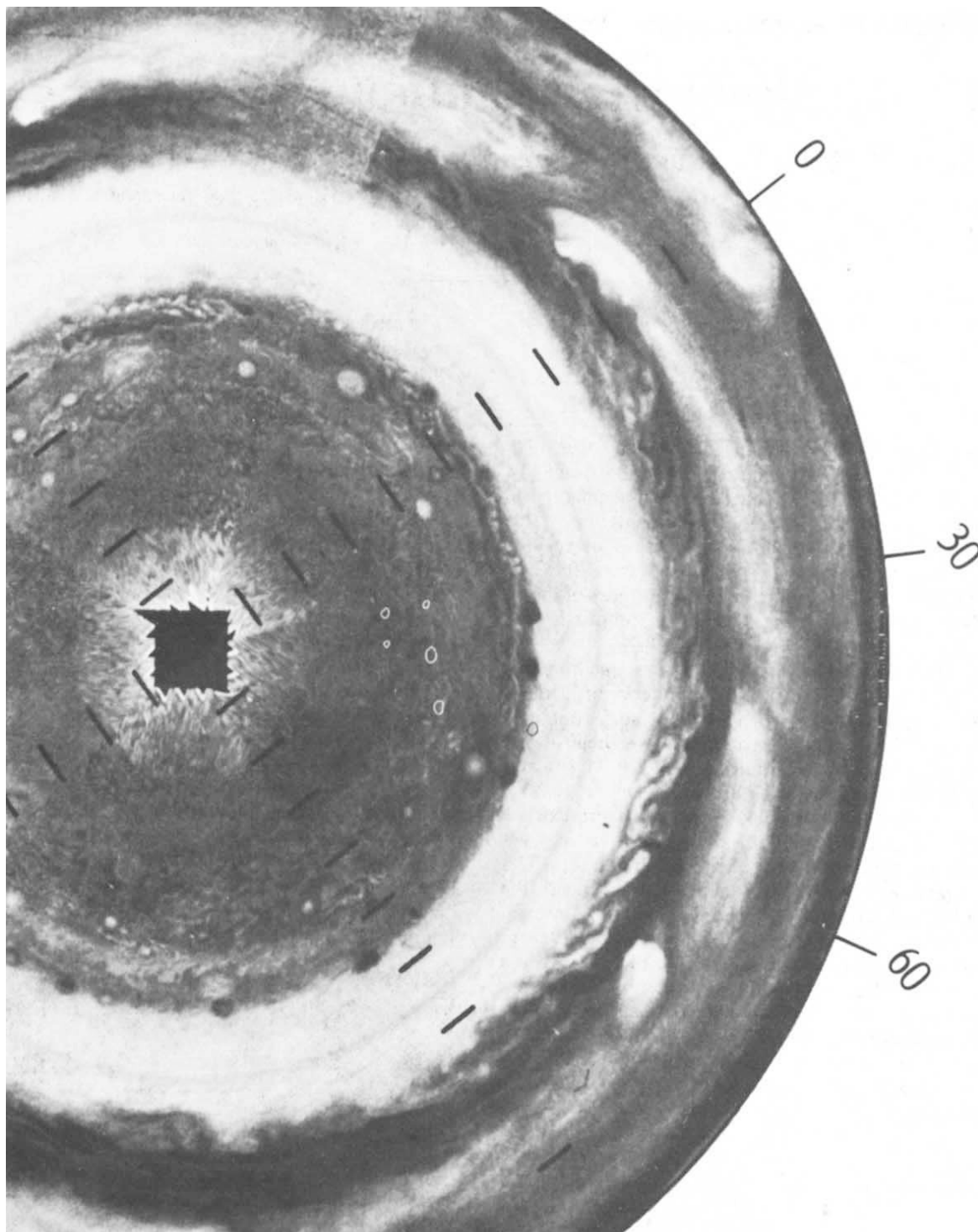
**Plate 13**

Plate 9 Portion of Jupiter from 25° to 145° W at 08.45 UT, 27 February 1979. This mosaic was made by projecting nine narrow angle frames in the orange filter onto a cylindrical grid. Distance along the ordinate is proportional to distance along the axis of rotation. Latitudes are planetographic, as described in the text. Resolution in the original frames is about 120 km, and is several times poorer when printed.

Plate 10 *a*, Textural map of the portion of Jupiter covered in the mosaic of *Plate 9* using the symbols defined in *Plate 8a*. Textural types generally run parallel to the large-scale bands, but are not uniquely defined by the bands. Linear textures tend to form chevron patterns centred on the zonal jet maxima and on the circumferential flow around the major spots. *b*, A flow diagram indicating observed jovian currents. Arrows indicate only the direction of flow, not its magnitude. The velocity field is overlaid on a cylindrically projected mosaic of nine Voyager narrow-angle frames taken through an orange filter on 27 February 1979.

Plate 11 A polar stereographic map of Jupiter's northern hemisphere made from 27 Voyager narrow-angle images taken over one jovian rotation on 1 February 1979. Longitudes are System III (1965.0), while latitudes are planetographic. Note the motion of cloud features relative to each other when compared with the regional mosaics (*Plate 10*) taken at a later date. Latitude tick marks are located at 10° intervals. The outer circle is located at 10° S. The Voyager 1 trajectory took the spacecraft on a near equatorial pass by the planet, hence data is missing near the pole.

Plate 12 A polar stereographic map of Jupiter's southern hemisphere made from 27 Voyager narrow-angle images taken on 1 February, 1979. Longitudes are System III (1965.0), while latitudes are planetographic. Again, note the differential motions of features when compared with *Plate 10*. Latitude tick marks are located at 10° intervals. The outer circle is located at 10° N. The south pole itself was unobservable in Voyager 1's equatorial plane encounter with Jupiter.

Plate 13 First Quadrant of a Polar Stereographic Projection of Jupiter's northern hemisphere on 1 February 1979 with the locations of the lightning flashes from Fig. 1 of Cook *et al.* (p. 794) superimposed. Latitudes are jovigraphic.

First results on jovian lightning

A. F. Cook II*, T. C. Duxbury† & G. E. Hunt‡

* Center for Astrophysics, Harvard College Observatory, and Smithsonian Astrophysical Observatory, Cambridge, Massachusetts 02138

† Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91103

‡ Laboratory for Planetary Atmospheres, Department of Physics and Astronomy, University College, London WC1, UK

The geometric reduction of the discovery picture of lightning on the dark side of Jupiter relates the positions of the lightning flashes to visible cloud structure.

THE observations of the dark side of Jupiter were taken at 1979 March 4 d 17 h 52 min 30 s UT (time at the spacecraft) [see Fig. 10 in ref. 1]. A total of 13 stars appeared as double images and were used to establish the direction and orientation of the camera. The lower bound of the aurora establishes the position of Jupiter's limb which matches that found from the best ephemeris of the spacecraft. Longitudes and latitudes of the 20 lightning flashes are plotted in Fig. 1, while adjacent flashes are grouped together.

These positions also appear plotted on the first quadrant of a polar stereographic projection centred on Jupiter's north pole in Plate 13, Fig. 2. This projection was made from images obtained on 1 February 1979 at the spacecraft. We believe the mechanisms of generation of lightning on Jupiter and Earth may be similar. The most plausible mechanism is convective electrification². Convection in the clouds distorts a pre-existing background charge distribution and thus collects the charges into a distribution which is discharged by lightning strokes. This mechanism requires a conductor some distance beneath the clouds. On Earth this is often the solid planet itself. However, the surface and a sufficiently conducting layer of cloud-free air will serve as well, as in the case of thunderstorms above warm fronts. This situation may also apply to the atmosphere of Venus in which lightning has also been detected³ and where the atmosphere is clear of clouds in the lowest 35 km. The extension of this concept to Jupiter requires a sufficiently conducting atmosphere below the observed clouds, which may include a cloud-free region.

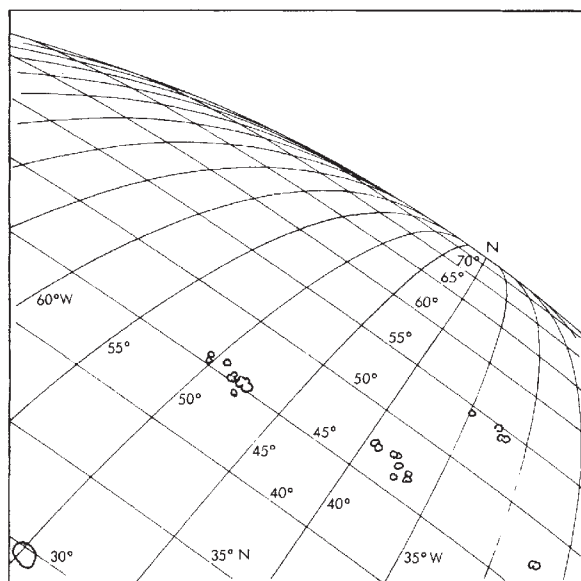


Fig. 1 The arrangement in the television field of Jupiter and the flashes of lightning. Latitudes are jovicentric.

The association of thunderstorms with upwelling in regional circulations is well known on Earth and also occurs on Jupiter. Consequently, observations of lightning may be used as a tracer of upwelling in the jovian circulation. The observations in Plate 13 indicate that these lightning flashes are concentrated in certain jovian latitudes, and atmospheric features which are located at the two lower latitudes ($+30^\circ$, $+46^\circ$ jovicentric; $+33^\circ$, $+50^\circ$ jovigraphic) are cloud structures in the form of bands between herringbone patterns of varying degrees of distinctness. In our case, at the lower latitude the lightning is associated with a northern branch of this band which can be double or triple but rejoined at another longitude. These bands are associated with unstable westward moving jets⁴. In the northern hemisphere (Fig. 2 in ref. 4) these jets occur at latitudes of 18° , 30° (with northward and southward branches) 36° , 45° , 50° and possibly by extrapolation, 54° and 58° N. At the time of these observations, lightning appeared at the second, fourth and sixth of these jets and not at the third and fifth which were also in view. This comparison leads us to suggest that lightning is a tracer of upwelling in the jovian atmosphere. The observation of these lightning flashes at high latitudes is consistent with the concept of internal heating generating convective cloud systems. Further support for this idea must await analysis of images from Voyager 1 and Voyager 2. Note that the detection of whistlers from Jupiter by the plasma wave instrument indicates the presence of lightning in the polar regions⁵.



Fig. 2 Locations of lightning flashes in Jupiter's northern hemisphere (from Plate 13, p. 793).

These results were made possible by all those who have worked in the Voyager project. Many of these have been individually acknowledged in ref. 1. We especially thank C. C. Avis, G. W. Garneau, P. L. Jepsen, J. J. Lorre, J. A. Mosher, G. M. Yagi, J. T. Harwood, L. Pieri and C. Reslock of JPL, and S. A. Collins as experiment representative at JPL. This paper reports one phase of research supported by NASA at JPL under contract NAS 7-100 and of research supported by JPL at the Smithsonian Astrophysical Observatory under that NASA contract. G.E.H. was supported by the SRC.

Received 19 July; accepted 31 July 1979.

1. Smith, B. A. *Science* **204**, 951-972 (1979).
2. Moore, C. B. & Vonnegut, B. in *Lightning* (ed. Golde, R. H.) 51-98 (Academic, London, 1979).
3. Taylor, W. W. L., Scarf, F. L., Russell, C. T. & Brace, L. H. *Nature* **279**, 614 (1979).
4. Ingersoll, A. P. et al. *Nature* **280**, 773-775 (1979).
5. Garnett, D. A., Shaw, R. R., Kirth, W. S. & Scarf, F. L. *Geophys. Res. Lett.* (submitted).

An interpretation of the Voyager measurement of jovian electron density profiles

S. K. Atreya, T. M. Donahue & J. H. Waite, Jr

Space Physics Research Laboratory, Department of Atmospheric and Oceanic Science, The University of Michigan, Ann Arbor, Michigan 48109

The large vertical extent and diurnal variation in the electron density profile of Jupiter observed in radio occultation by Voyager 1 imply a homopause eddy diffusion coefficient of $1\text{--}3 \times 10^5 \text{ cm}^2 \text{ s}^{-1}$ and an exospheric temperature of about 1,300 K.

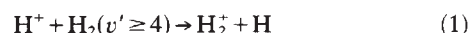
THE Voyager 1 radio science experiment (RSS) obtained electron density profiles for the daytime and night-time ionosphere of Jupiter by radio occultation observations on 5 March 1979¹. The properties of the ionosphere were different from those obtained by the Pioneer 10 and 11 observations^{2,3} in December 1973 and December 1974, in several important ways. (1) The vertical extent of the ionosphere in 1979 was about 6,000 km, compared to 3,500–4,000 km at the time of the Pioneer passages. (2) The peak in the daytime electron density occurred about 1,600 km above the reference pressure level (1 mbar) in 1979, compared to ~1,000 km above that level in 1973–74. (3) The scale height of the topside ionosphere in the daytime changes from about 590 km just above the peak to about 960 km at higher altitudes. In the night-time, there is a single topside scale height of 960 km corresponding to the high altitude daytime scale height. (4) Finally, the maximum electron density is a factor of 10 lower on the nightside than it is on the dayside. In the case of the Pioneer observations, only small variations were observed between daytime and night-time ionospheric profiles, and very small differences were evident as a function of latitude (20° N, 26° N, 58° N, and 79° S)³. The Pioneer 10 and 11 encounters with Jupiter occurred near solar minimum, while the Voyager 1 encounter was near the solar maximum. Here we attempt to account for the differences between the Pioneer and Voyager observations on the basis of solar cycle changes in solar flux, drastic decrease in the eddy diffusion coefficient in the upper atmosphere and a large increase in the exospheric temperature of Jupiter.

Atreya and Donahue⁴ succeeded in accounting for the principal features in the Pioneer 10 and 11 ionospheric profiles by assuming a temperature of 1,000 K for the upper atmosphere and invoking radiative recombination of H^+ as the dominant ion loss mechanism in the principal ionospheric layer. Here we argue, first, that the temperature has increased to 1,300 K in 1979. This neutral temperature corresponds to the plasma scale height of 960 km which was measured at 3,500 km above the reference level on the dayside and at 2,300 km on the nightside. A lower temperature (1,100 K) was quoted by Eshleman *et al.*¹ This was the consequence of using the plasma scale height at 6,000 km. (G. F. Lindal, personal communication). However, photochemical time constants are small compared to diffusion times only below about 3,500 km, and, hence, we have chosen to calculate the temperatures at lower altitudes. Nagy *et al.*⁵ have shown that approximate thermal equilibrium between the plasma and neutral gas occurs at this altitude. The difference between neutral, electron and ion temperature should be at most a few hundred degrees and has only a small effect on our results.

We have varied the eddy diffusion coefficient at the homopause from $3 \times 10^7 \text{ cm}^2 \text{ s}^{-1}$ which was appropriate for Pioneer 10

and 11 conditions⁴ to $8 \times 10^4 \text{ cm}^2 \text{ s}^{-1}$ in order to ascertain the effect of a change in that atmospheric parameter on the ion density profiles.

We have also taken into account the variation with temperature of rate constants. The rate constant for radiative recombination of H^+ and electrons varies with electron temperature^{6,7} between $T_e^{-0.5}$ and $T_e^{-0.75}$. A reaction discussed by McElroy⁸



becomes much more important at Voyager temperatures than those of Pioneer 10 and 11. Followed by



Reaction 1 can be a very significant loss process for H^+ ions in the high temperature conditions encountered by Voyager.

Finally, extreme UV solar flux values appropriate to the conditions of the Voyager encounter were taken from those measured by Hinteregger (personal communication).

We have found that it is possible to reproduce the measured electron density profiles by using a model atmosphere appropriate to an exospheric temperature of 1,300 K with temperature varying above the homopause as in our previous calculations⁴ and with an eddy diffusion coefficient K of $(1\text{--}3) \times 10^5 \text{ cm}^2 \text{ s}^{-1}$ at the homopause. K varies with altitude inversely as the square root of the atmospheric number density above the tropopause. Its value at the tropopause was adjusted until an electron density profile matching the observed one was obtained. For example, in a model in which the value of the H_2 density at the homopause was $7 \times 10^{13} \text{ cm}^{-3}$ and K at that altitude was $1.5 \times 10^5 \text{ cm}^2 \text{ s}^{-1}$, the daytime maximum electron density of $2 \times 10^5 \text{ cm}^{-3}$ occurred where the H_2 density was $3 \times 10^{11} \text{ cm}^{-3}$, 1,500 km above the 1-mbar level. (The observed peak was at 1,600 km.) The hydrogen density at the ionospheric maximum is somewhat higher than it was for the conditions of the Pioneer encounters⁴ when the thermosphere was cooler and the solar flux was weaker. Given the present state of uncertainty in measured altitudes and plasma profiles, it is probably safe to take K as lying somewhere between 1 and $3 \times 10^5 \text{ cm}^2 \text{ s}^{-1}$ at the homopause.

The rate of ionisation during the daytime is 26 protons $\text{cm}^{-3} \text{ s}^{-1}$ at the maximum. This is a full order of magnitude above the rate required to account for the ionosphere observed by the Pioneers. It is partly the consequence of the 2.5-fold increase in ionising solar flux that has occurred since the last solar minimum (H. E. Hinteregger, personal communication). It is also a result of the fact that an increase in atomic hydrogen column abundance in the upper atmosphere of Jupiter by a factor of 100 since 1973–74 is implied by the very large enhancement of jovian Ly α radiation detected by the UVS⁹ and IUE (H. W. Moos and S.K.A. unpublished) observations at the time of Voyager encounter.

The time constant for loss of H^+ by radiative recombination is substantially greater than the length of a jovian night. This circumstance led to the small diurnal variations observed by Pioneers 10 and 11⁴. We propose that the diurnal variation now

observed leading to the tenfold enhancement in the maximum electron density, and the dual topside scale height during the day is a consequence of the enhanced importance of reaction (1) when the temperature in the upper atmosphere is 1,300 K in place of 850–1,000 K. In order to reproduce the observed diurnal variation, the loss frequency for H^+ in charge exchange with H_2 should be $1.3 \times 10^{-4} s^{-1}$, leading to an overall rate constant for reaction (1) of $4.3 \times 10^{-16} cm^3 s^{-1}$. Reaction (1) is endothermic in the ground vibrational state by 1.8 eV for H^+ at 300 K¹⁰. The rate constant for H_2 excited to the fourth vibrational level and higher at a temperature of 1,300 K would have to be $4 \times 10^{-9} cm^3 s^{-1}$ to produce the required overall loss rate. This value would be reduced by about a factor of two if we take into account the high velocity of the H^+ ions at 1,300 K. Because of the rapid variation of the Boltzmann factor with temperature, a small increase in vibrational temperature would also lead to a sizeable relaxation in the value required for this rate constant. On the other hand, if the temperature should be 850 K as it apparently was during the Pioneer 10 and 11 observations, the rate constant for the reaction between H^+ and H_2 excited levels higher than $v = 4$ would have to be of the order of $10^{-5} cm^3 s^{-1}$ to produce a similar reduction in ion density.

The photochemical time constants are short compared with a jovian day up to an altitude where the dayside electron density is about $2 \times 10^4 cm^{-3}$. Hence, after sunset the electron density near

the daytime peak will rapidly decay. The night-time ionospheric maximum will attain a density of $\sim 10^4 cm^{-3}$ and a single topside scale height corresponding to the upper part of the daytime profile. The nightside ionosphere is the dayside ionosphere with its low altitude bulge dissolved.

The ionosphere observed by Voyager 1 seems to imply the existence of a jovian exospheric temperature at 1,300 K and an eddy diffusion coefficient between 1 and $3 \times 10^5 cm^2 s^{-1}$ at the homopause. This high temperature and low value of K compared with Pioneer conditions are also implied by the great enhancement in $Ly\alpha$ radiation observed by the Voyager UVS, and IUE, and the fact that the UVS failed to detect the He 584 Å radiation from the jovian disk. If the volume mixing ratio of He is 0.11 as reported by the IR spectroscopy and radiometry investigation¹¹, the failure of the helium resonance radiation to be observed by the UVS implies the presence of a large amount of hydrogen above the homopause and a value for K near the homopause of the order of $10^5 cm^2 s^{-1}$. Although the rate constant we have calculated for reaction (1) seems reasonable, it will be highly desirable to measure it in the laboratory in jovian conditions.

We thank Von R. Eshleman and G. Fjeldbo Lindal for valuable discussions. This research was supported by NASA grant NSG 7404 and the Atmospheric Sciences Section of the NSF.

Received 27 June; accepted 11 July 1979.

1. Eshleman, V. R. *et al. Science*, **204**, 976–978 (1979).
2. Fjeldbo, G. *et al. Astr. Astrophys.* **39**, 91 (1975).
3. Fjeldbo, G. *et al. Jupiter* (ed. Gehrels, T.) 238 (The University of Arizona Press, 1976).
4. Atreya, S. K. & Donahue, T. M. *Jupiter* (ed. Gehrels, T.) 304 (The University of Arizona Press, 1976).

5. Nagy, A. F. *et al. J. geophys. Res.* **81**, 5567–5569 (1976).
6. Bates, D. R. & Dalgarno, A. *Electronic Recombination, Atomic and Molecular Processes* (ed. Bates, D. R.) 245 (Academic, New York, 1962).
7. Bauer, S. J. *Physics of Planetary Ionospheres*, 84 (Springer, New York, 1973).
8. McElroy, M. B. *Space Sci. Rev.* **14**, 460 (1973).
9. Broadfoot, A. L. *et al. Science* **204**, 979–982 (1979).
10. Browning, R. & Fryar, J. *Proc. Phys. Soc. Lond. at. molec. Phys.* **6**, 364–371 (1973).
11. Hanel, R. A. *et al. Science* **204**, 972–976 (1979).

Plasma wave turbulence at Jupiter's bow shock

F. L. Scarf*, D. A. Gurnett†, W. S. Kurth† & R. L. Poynter‡

* Space Sciences Department, TRW Defense and Space Systems Group, Redondo Beach, California 90278

† Department of Physics and Astronomy, The University of Iowa, Iowa City, Iowa 52242

‡ Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91103

The first measurements of the wave-particle interactions of Jupiter's bow shock are reported, and some of the wave phenomena detected during the inbound passage are discussed.

THE characteristics of a planetary bow shock are important because the plasma that impacts the magnetosphere of a planet represents solar wind that has been drastically modified by the shock process. Plasma instabilities that develop at a shock front are associated with significant acceleration processes that yield non-maxwellian distributions in the downstream magnetosheath as well as significant fluxes of suprathermal ions and electrons in the upstream region. The Voyager 1 plasma wave investigation has provided the first opportunity to study directly the wave-particle interactions of an outer planet bow shock. Here we discuss in detail some of the wave phenomena detected during the inbound passage.

Measurements

The initial report on the Voyager wave data¹ contains a brief summary of the inbound shock measurements with a display of the 16-channel spectrum analyser data from the first bow shock crossing (detected at 14.34 spacecraft time, 28 February 1979,

when the spacecraft was at a range of 85 R_J). Figure 1 shows a corresponding plot of the 16-channel wave data for the third shock crossing of 12.27 UT on 1 March, detected when Voyager 1 was at a range of 72 R_J . We focus attention here on this third inbound crossing for which we have the most comprehensive data records.

Before 12.27, Fig. 1 shows that electrons produced steady oscillations in the 5.62 kHz channel, corresponding to an upstream density of ~ 0.4 electrons cm^{-3} . The high frequency upstream wave amplitudes were almost an order of magnitude higher than those detected in the corresponding region the day before, but in the low-frequency channels where ion acoustic waves and whistlers are generally detected, the 1 March bow shock exhibited a noteworthy absence of precursor wave activity. (The sporadic signals in the 100-Hz channel are interference tones from the stepper motor of the low energy charged particle instrument.) This third shock crossing was characterised by the detection of intense broadband noise enhancements in a well defined time interval with duration of the order of 1 min. Analyses of collisionless shock structures² suggest that the minimum shock thickness for a laminar structure should be of the order of $\delta_i = c/(2\pi f_p)$ where c is the speed of light, f_p is the electron plasma frequency ($f_p = 9\sqrt{N_e}$ kHz where N_e is the electron density in cm^{-3}) and $f_p = f_p/43$ (for protons). For our

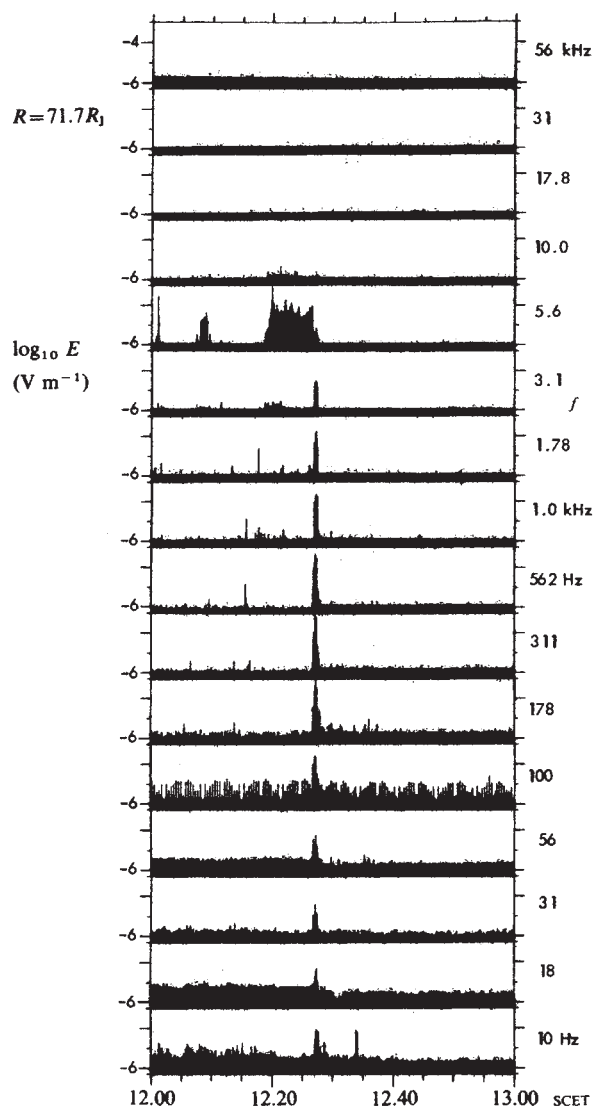


Fig. 1 Measurements of wave amplitudes at the third bow shock crossing. Interference appears in the 100 Hz channel, and the upstream enhancements at 5.6 kHz represents detection of electron plasma oscillations. Note the very low noise levels in the magnetosheath after 12.27.

case we know that $f_p^* = 5.6$ kHz, and hence $\delta_e (= \delta_i/43)$ should be ~ 8.5 km while $\delta_i \sim 365$ km. At this time the Voyager radial speed with respect to the planet centre was ~ 12.7 km s $^{-1}$, and hence the spacecraft would have traversed a stationary minimum shock with thickness of 8.5 km in ~ 2 s. The observed duration of about 20–60 s seems to imply that the thickness corresponded to an ion inertial length.

At the time of this shock crossing we were fortunate to have high rate waveform data available that provided complete spectral information over the range 50 to 12 kHz. Figure 2 shows the frequency–time plot made up using the high rate data from 1226.12 to 1227.48 spacecraft time. (A Voyager-to-Earth data transmission of just over 11 Mbits was used to provide the

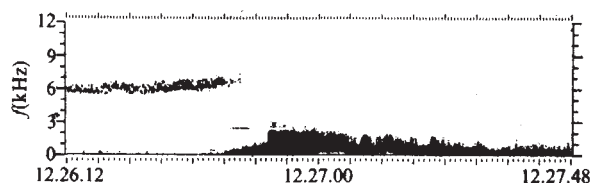


Fig. 2 Frequency–time diagram of bow shock number three made up using the Voyager 1 high rate data link. The upstream electron plasma oscillation line is highly structured just before the shock encounter.

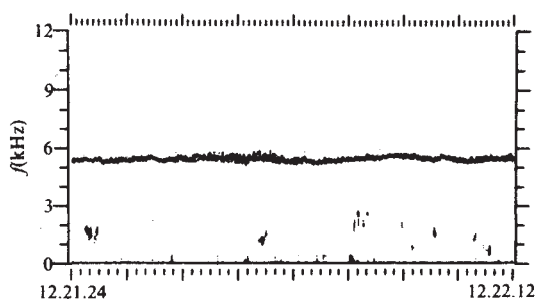


Fig. 3 Frequency–time diagram of regular upstream electron plasma oscillations detected a few minutes before the shock.

plasma wave information in Fig. 2.) The thick structured 'line' with $f \approx 6$ kHz (before 12.26.45) represents the upstream electron plasma oscillation emission, and we identify the low frequency wave turbulence after 12.26.42 as ion acoustic waves. The upstream electron emission just before this shock has a distinct and unusual structure that may be related to a nonlinear effect operative very near to the shock. Figure 3 shows a frequency–time plot made up just a few minutes earlier with very narrow and unstructured electron plasma oscillation lines, and it is possible that the structure shown in Fig. 2 is related to the coupling of large amplitude electron plasma oscillations and ion acoustic waves³. Similar phenomena have occasionally been detected in the region immediately upstream from the Earth's bow shock⁴. However, as Fig. 2 does not show enhanced ion sound wave levels just before 12.26.42, other explanations for the high frequency wave structure may have to be considered. On the other hand, the broadband channel uses an automatic gain control amplifier that responds primarily to the most intense waves, and the bandpass channel data (Fig. 1) show that the upstream ion waves in the Fig. 2 time interval were at least as intense as the ones shown in Fig. 3.

Figure 2 shows several other features of interest. First we note that the plasma wave activity was very low by the end of the record. This result is also indicated in Fig. 1, which shows that the jovian magnetosheath is characterised by a virtual absence of detectable plasma wave turbulence after passage through the shock. (Sarf *et al.*¹ noted that this is true for the entire sheath traversal into the magnetosphere boundary.) Figure 2 also shows a number of impulsive wave structures within the shock with durations as small as 1, 2 or 3 s (see especially the period after 12.27.06) and it is possible that these structures represent regions with scale sizes comparable to $\delta_e = c/(2\pi f_p^*)$.

Clearly the high frequency and time resolution available from the broadband data link will allow us to carry out detailed studies of the thermalisation and acceleration processes that develop at the very strong jovian bow shocks, and future studies will focus attention on the shock structure and on correlations with data from other Voyager instruments. Recent detailed calculations of shock structure based on the use of marginal stability criteria⁵ will be applied here when the complete plasma and magnetic field profiles are available. However, it is already apparent that Jupiter's bow shock region provides material for novel and significant plasma physics studies and future research will also be aimed at understanding the absence of magnetosheath noise, as well as the structured electron plasma emissions detected just upstream from the shock.

The research at TRW was supported by NASA contract 954012 with the Jet Propulsion Laboratory. Work at the University of Iowa was supported by NASA contract 954013 with JPL and by grant NGL-16-001-043 with NASA headquarters. Research at JPL was supported by NASA contract NAS7-100.

Received 16 July; accepted 31 July 1979.

1. Sarf, F. L., Gurnett, D. A. & Kurth, W. S. *Science* **204**, 991–995 (1979).
2. Greenstadt, E. W. & Fredricks, R. W. *Magnetospheric Physics* (ed. McCormac, B. M.) 281–290 (Reidel, Dordrecht 1974).
3. Smith, R. A., Goldstein, M. L. & Papadopoulos, K. *Astrophys. J.* (in the press).
4. Anderson, R. R., Gurnett, D. A. & Sarf, F. L. *Eos* **59**, 1156 (1978).
5. Morse, D. L. & Greenstadt, E. W. *J. geophys. Res.* **81**, 1791–1793 (1976).

In situ identification of various ionic species in Jupiter's magnetosphere

James D. Sullivan & F. Bagenal

Center for Space Research, Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Positive ions with various mass per charge values up to ~160 have been identified within 20 jovian radii of Jupiter from the analysis of data from the plasma science experiment on Voyager 1.

ONE of the major surprises from the Voyager 1 encounter with Jupiter in March 1979 was the discovery¹ of active volcanoes on Io. This result emphasises the importance of Io as the principal source of the bulk plasma in the jovian magnetosphere^{2,3}. The plasma science experiment⁴ on Voyager 1 has made the first detailed *in situ* measurements of the plasma (10–5,950 V) near Jupiter². Continuing analysis of these measurements for the ionic composition of this plasma has revealed the existence of further atomic and molecular ions as minor constituents of the plasma. We have now identified ions with mass per charge (A/Z^*) values of 1, 8, 10–2/3, 16, 23, 32, 64, ~104 and ~160 within 20 jovian radii, R_J , of Jupiter in the dayside magnetosphere.

The measurements used in the preliminary analysis reported here were taken in the high resolution mode (M mode) which has ~3.6% resolution in energy per charge between 10 and 5,950 V. The energy per charge scan effectively becomes a mass per charge scan whenever all ionic species have a common component of velocity into any sensor; this common component of velocity as well as the spacecraft potential are parameters of the present analysis. (The instrument contains four independent sensors⁴ A, B, C and D, looking in different directions⁵.) The measurements are usually displayed as relative distribution function versus energy per charge; the relative distribution function is the measured current divided by the energy per charge width of the particular measurement step. Since the step widths increase with increasing energy per charge, a constant background will result in a smaller value of the distribution function at higher values of energy per charge.

Significant heavy ion abundances are found beyond the inner magnetosphere where ions with $A/Z^* = 8, 16, 32$ and 64 were reported by Bridge *et al.*². For example, at 19.6 R_J , the energy/charge spectrum (Fig. 1) shows several distinct peaks.

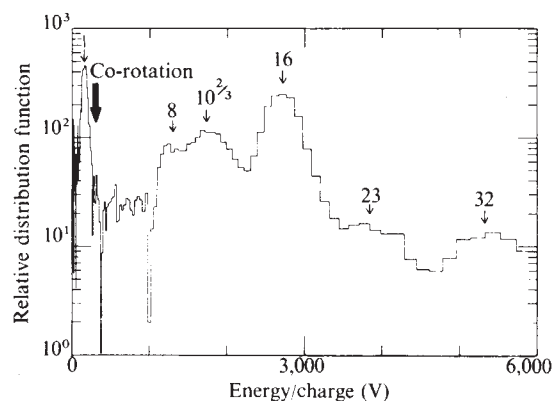


Fig. 1 D-sensor energy per charge spectrum obtained on day 63 at 15.50 UT when the spacecraft was at 19.6 R_J and a magnetic latitude of -8.5° . The heavy arrow labelled co-rotation marks the expected location of a proton peak moving with the geometrically expected co-rotation velocity. The light arrows mark the locations of various A/Z^* values assuming the lowest peak is a proton peak.

This spectrum is typical of the spectra for ~15 min about this time. Each spectrum is analysed with a simultaneous fit to a sum of convected isotropic maxwellian distribution functions with a common component of velocity; the sum is over peaks at $A/Z^* = 1, 8, 10-2/3, 16, 23$ and 32. The analysis of this spectrum gives a common component of velocity of $\approx 180 \text{ km s}^{-1}$ and a spacecraft potential of less than $\pm 5 \text{ V}$. The common component of velocity is not identical to the value of $\approx 240 \text{ km s}^{-1}$ expected geometrically from co-rotation; this difference and related measurements are discussed by McNutt *et al.*⁵. The mass/charge values for the individual peaks obtained from the fit are within half a unit of the indicated numbers. Thus, we can make the following tentative assignments: $A/Z^* = 1$ is H^+ ; 8 is O^{2+} or S^{4+} ; 10–2/3 is S^{3+} or even B^+ ; 16 is O^+ or S^{2+} ; 23 is Na^+ ; and 32 is S^+ (or O_2^+). The peak observed at $\approx 3.9 \text{ kV}$ is most probably sodium and not magnesium or neon. This conclusion is based on the temperatures of the ions derived from the maxwellian fits; within $\pm 20\%$ all ionic species have the same temperature. Under these circumstances, it is possible to produce a peak at $A/Z^* = 23$ by an appropriate combination of Mg^+ and Ne^+ . However, the temperature of the 'sodium' peak produced in this way is too high in comparison with temperatures of the other ions. There is a clear absence of a He^+ or He^{2+} peak; a rough upper limit can be put at $n(\text{He}^+, \text{He}^{2+}) < \frac{1}{15}n(\text{H}^+)$. Any signal from $A/Z^* = 64$ would come in at $\approx 10.8 \text{ kV}$ well above our energy per charge scan range. The mass ratio of heavy ions to protons can be computed independently of the actual

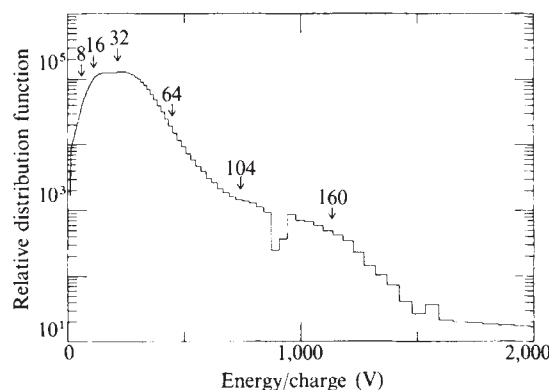


Fig. 2 C sensor energy per charge spectrum obtained on day 64 at 10.00 UT when the spacecraft was at 5.4 R_J and a magnetic latitude of $+4.4^\circ$. The light arrows mark the locations of various A/Z^* values assuming strict co-rotation. The notch at $\approx 950 \text{ V}$ is caused by interference from another experiment on the spacecraft.

mass and charge of the individual peaks as it depends on $(A/Z^*)^{3/2}$. The heavy ion to proton mass ratio is >95 ; consequently the mass density is dominated by heavy ions. It is instructive to compare these results with the inner magnetosphere results of ref. 2. First, the $A/Z^* = 23$ peak would have been masked by the prominent $A/Z^* = 32$ peak so that its absence at 5.3 R_J is not significant. Second, the relative abundances of 32:16:10–2/3 differ markedly; the higher ionisation states are enhanced at the earlier time, greater distance, reported here.

Direct measurement of values of $A/Z^* > 32$ were restricted to regions where the densities are high and the common component of velocity is small. Such is the case in the Io plasma

torus where the instrument view directions were favourable for the A, B and C sensors^{2,5}. Throughout this time protons moving with the geometric co-rotation velocity are below the energy/charge range of the instrument. Typical of the spectra in the torus is the energy/charge spectrum in Fig. 2; this spectrum is from the C sensor but the same results are obtained from the other two sensors (A and B). Compared with Fig. 1, the instrument gain has been reduced by a factor of ~ 77 in order to avoid saturation. Thus the detection threshold is raised and the curve above $\sim 1,600$ V corresponds to the minimum detectable signal; consequently the graph only extends to 2 kV. The purpose of this figure is to establish the existence of detectable intensities of high mass/charge ions. The signal between ~ 500 and $\sim 1,400$ V shows a broad shoulder well above instrument threshold. We find acceptable fits to this spectrum including peaks at $A/Z^* = 104 \pm 5$ and 160 ± 10 moving with a common component of velocity equal to the co-rotation value; a clean peak at $A/Z^* = 64$ was shown in Fig. 5 of ref. 2, the spectrum at 10.15 UT. The spectrum at 10.15 UT also has a peak at 104 ± 5 . These heavy ion peaks are almost certainly molecular ions. The $A/Z^* = 64$ peak could well be either SO_2^+ or S_2^+ . The $A/Z^* = 104$ peak is not consistent within the error with ZnS^+ at 96. Speculative

assignment for the $A/Z^* = 104 \pm 5$ and $A/Z^* = 160 \pm 10$ peaks may be made based on ions from sulphur salts or polymers which are suggested by observations of the surface of Io.

To conclude: inside $20 R_J$: (1) The plasma is composed of ions with mass to charge ratios 1, 8, $10\text{--}2/3$, 16, 23, 32, 64, ~ 104 , ~ 160 . (2) Wherever both protons and heavy ions are detected, the mass density is dominated by the heavy ions by a factor of ~ 100 . (3) The plasma ions move with a common component of velocity which is not always the value expected geometrically from co-rotation. (4) The ions with $A/Z^* \geq 64$ are probably molecular ions.

We thank J. W. Belcher, H. S. Bridge, A. J. Lazarus and R. L. McNutt for helpful comments. This work was supported by NASA contract 953733.

Received 27 June; accepted 24 July 1979.

1. Morabito, L. A., Synnott, S. P., Kupferman, P. N. & Collins, S. A. *Science* **204**, 972 (1979).
2. Bridge, H. S. *et al.* *Science* **204**, 987–991 (1979).
3. Broadfoot, A. L. *et al.* *Science* **204**, 979–982 (1979).
4. Bridge, H. S. *et al.* *Space Sci. Rev.* **21**, 259–287 (1977).
5. McNutt, R. L., Belcher, J. W., Sullivan, J. D., Bagenal, F. & Bridge, H. S. *Nature* **280**, 803 (1979).

Jupiter's magnetic tail

N. F. Ness*, M. H. Acuna*, R. P. Lepping*, K. W. Behannon*,
L. F. Burlaga* & F. M. Neubauer†

* Laboratory for Extraterrestrial Physics, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771

† Technische Universität, Braunschweig, FRG.

Magnetic field observations of the jovian magnetosphere suggest an extended magnetic tail, which has been formed by solar wind interaction with the planetary field.

In March 1979, Voyager 1 became the third spacecraft to penetrate and study *in situ* the magnetosphere of Jupiter. Earlier observations from Pioneers 10 and 11 in 1973–74 indicated the development of a magnetodisk topology describing the magnetosphere^{1,2}. In this model, the combined effects of rapid rotation and a strong planetary magnetic field yield an equatorial region with a considerably enhanced charged particle and plasma population and distended magnetic field lines³. Thus, the central region of the magnetodisk effectively carries an azimuthal electrical current and this region is commonly referred to as the current sheet. Controversy arose over whether the current sheet of the magnetodisk was planar and parallel to the magnetic equatorial plane, with a spiraling of the magnetic meridian planes and field lines due to a finite Alfvén speed^{4,5}, or if the current sheet and magnetodisk were distorted due to the centrifugal forces and deviated or warped so as to become parallel to the jovigraphic equatorial plane at distances greater than $20 R_J$ (ref. 6). An attempt was also made to interpret the data qualitatively in the framework of a 'magnetic anomaly' model⁷, based on the highly asymmetric planetary field^{6,8}, which did not include an azimuthal current sheet. All these earlier interpretations were based on a planet-centred view of the processes and mechanisms which control the configuration of the outer magnetosphere. We discuss here the Voyager I experimental observations which are most naturally interpreted in terms of a well developed magnetic tail on the nightside of the jovian magnetosphere. This tail, with a 'neutral sheet' separating the upper and lower lobes of opposite field polarity, is formed and controlled by the external forces associated with the solar wind interaction. The inner magnetosphere's current sheet is found to merge with the tail's neutral sheet. This configuration

leads to a strong local time control of the outer jovian magnetosphere configuration, in contrast to earlier models which emphasised inner, that is planetary, control.

The preliminary results obtained from the Voyager 1 magnetometer have already been summarised⁷. The Voyager 1 spacecraft entered the jovian magnetosphere at 11.00 LT with magnetopause crossings observed at jovicentric distances between 67 and $47 R_J$ and left the magnetosphere at 04.00 LT with magnetopause crossings occurring between 153 and $170 R_J$. Characteristic current sheet-neutral sheet signatures were observed in the magnetic field data, as the spacecraft passed through the magnetosphere. They showed up as significant decreases in the magnetic field intensity, occurring simultaneously with variations in the direction of the magnetic field. These are interpreted in the inner magnetosphere, $R < 30 R_J$, as evidence for a large scale azimuthal current or a current sheet within the magnetosphere.

Within $30 R_J$, these magnetically characteristic regions occurred nearly simultaneously with the crossing of the magnetic equatorial plane. Although the field intensity decreased appreciably, it never reached zero, and the variation of field direction was gradual and not consistent with the nearly anti-parallel field directions on opposite sides of the current sheet which were observed at larger radial distances. Beyond $30 R_J$, the characteristics of the current sheet crossings were similar to those seen when crossing the neutral sheet in the Earth's magnetic tail: a sharply defined dip in the field intensity, to very small values, while the field direction changed by approximately 180° . Within $80 R_J$, there were two sheet crossings every 10 h, although not at a 5 h spacing. Beyond $80 R_J$, complete traversals of the current sheet were not observed as evidenced in the field direction. However, very close approaches were made, as seen in the field intensity dips and increased higher frequency fluctuations which occurred with regularity at the 10 h rotation period of the planet. The times of the occurrence of the full traversals merged as the two traversals per rotation period changed to one partial traversal per rotation period.

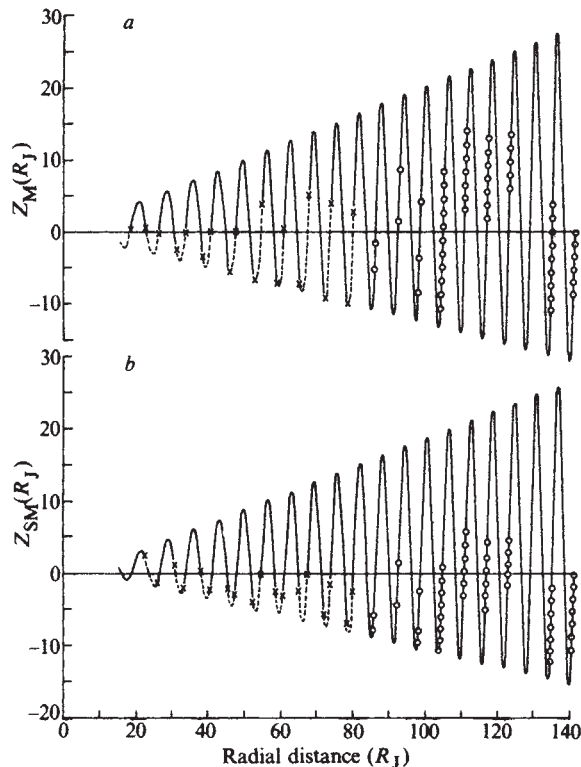


Fig. 1 Position of Voyager 1 spacecraft during exit from Jovian magnetosphere in magnetic dipole (a) and solar magnetospheric (SM-b) coordinates. SM coordinates are right-handed, non-rotating and defined with x_{SM} axis from planet to sun, z_{SM} axis in plane formed by x_{SM} and M , the magnetic dipole moment of the planetary field. Thus, they rock back and forth about the planet sunline following the magnetic dipole. — Above NS; --- below NS; ×, NS; ○, depressed field.

Figure 1 shows the trajectory of the spacecraft outbound from periaapsis as observed in both jovian dipole magnetic coordinates (tilt = 9.6° and System III (1965) long = 202°) and solar magnetospheric coordinates. This shows the change from planetocentric control of the occurrence of sheet crossings to solar wind control in the outer magnetosphere. Figure 1a shows that within $25 R_J$, the current sheet crossings, denoted by crosses, clearly occur nearly coincident with the spacecraft traversal of the magnetic equatorial plane. Contrast this with Fig. 1b in which there is no consistency with traversal of the solar magnetospheric equatorial plane. In the outer portion of the Voyager 1 magnetosphere traversal, the magnetic dips, indicated by open circles, are seen to occur during those portions of the trajectory which were near or below the solar magnetospheric equatorial plane. The dips were sometimes observed as extended periods of reduced and variable field strength. Such intervals are shown by sequences of open circles. The conclusion from this figure is that the magnetospheric structure is characterised by a transition from a magnetic equatorial current sheet to the neutral sheet of the tail which is approximately parallel to the solar magnetospheric X-Y plane.

Earlier studies of the Pioneer 10 outbound trajectory had primarily emphasised the times of crossing or penetration of the current sheet^{2,3}. In order to understand the magnetic field observation correctly, however, it is also important to consider properly the configuration of the magnetic field. In order to study these Voyager 1 data further, we transform the vector observations into solar magnetosphere coordinates. Figure 2 presents a projection of the hourly averaged X-Y and Y-Z components in both the X-Y and Y-Z planes along the trajectory of Voyager 1. The position of the magnetopause is included in the Fig. 2a and was derived by matching the mid

points of the inbound and outbound regions of multiple crossings. The length of the field vectors was scaled logarithmically as $K(1 + \log B_{ij})$ where $ij = xy$ or yz , with representative values of 1 and 100 nT illustrated. The periodic traversal of the current sheet within $80 R_J$ is immediately evident in the alternating direction of vectors in the X-Y projection. Those vectors with a positive X component were observed above the current sheet and a negative X component below. Note that the field direction observed in the X-Y plane does not change significantly along the trajectory and indeed approaches a direction, which as the magnetopause is approached, is parallel to the magnetopause.

The X-Z projection shows that close to the planet there is a substantial negative Z component, consistent with the inner magnetosphere being dominated by the planetary dipole term. The field vectors beyond $30 R_J$ tend to be more parallel to the solar magnetospheric equatorial plane for the most part. However, the field vectors do show an increasingly southwards tilt at larger radial distances and at increasingly smaller and also more negative values of the spacecraft z_{SM} position.

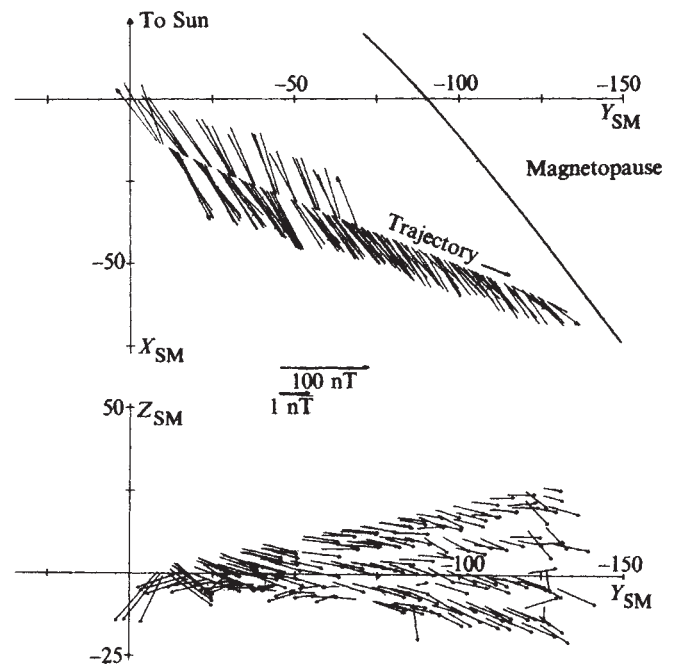


Fig. 2 Projection of hourly average magnetic field components, in solar magnetosphere coordinates, along Voyager 1 trajectory. The observed stratification in the lower panel into four apparently distinct levels is a consequence of the relative phasing between the hourly averages and the very nearly 10 hour rocking periodicity of the coordinate system.

Comparing these data with Fig. 1 in SM coordinates, it is noted that both the orientation of the field at negative SM latitudes and the occurrence of incomplete current sheet crossings is consistent with a current sheet merging with a neutral sheet surface which bends southwards at larger distances at this time. Due to the long transit time from periaapsis to first outbound magnetopause crossing, 10 days, some of the variations in the data may reflect temporal variations in the magnetosphere structure. These are convolved with large scale spatial variations.

The direction of the field and the position of the current sheet and neutral sheet, as either directly observed or inferred, is reminiscent of observations made near Earth by spacecraft traversing the magnetic tail in the dawn region of the terrestrial magnetosphere. Figure 3 presents magnetic field observations from the IMP 1 satellite which discovered the Earth's magnetic tail¹⁰. Figure 3 shows that near the dawn terminator, the field is distorted from the magnetic meridian planes and is bent back-

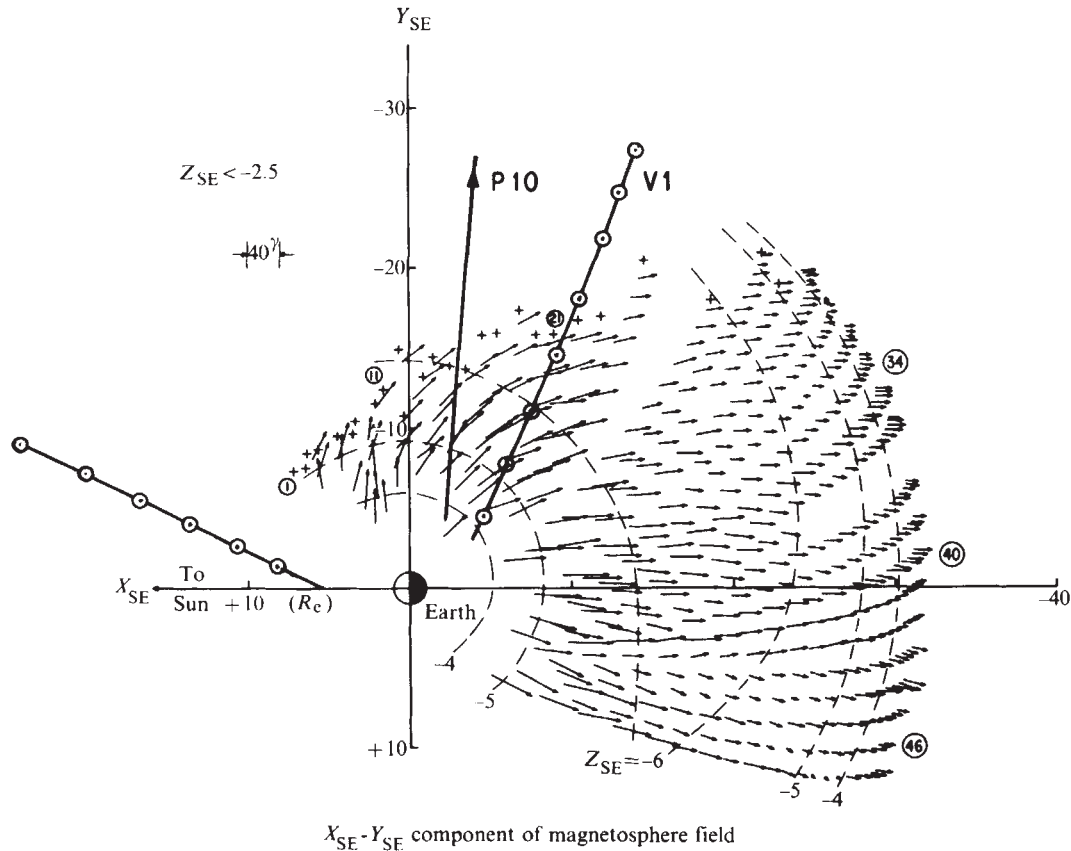


Fig. 3 Relative position of Voyager 1 and Pioneer 10 trajectories superimposed on results obtained by a survey of the Earth's magnetic tail by IMP-1 (Explorer 18). +, Observed magnetopause position.

wards so as to parallel the magnetopause boundary. Superimposed on this diagram are the trajectories of Voyager 1 out-bound and Pioneer 10 assuming equal scaling of the positions of the magnetopauses. The Voyager 1 magnetic field results as shown in Fig. 2a are consistent with those to be expected from this figure. Figure 3 also shows evidence of a magnetic tail along the Pioneer 10 trajectory less tailwards than Voyager 1. The field orientation would change from the magnetic meridian plane when close to the planet, and become gradually deflected tailwards until an angle of approximately 60° is reached near the dawn terminator, paralleling the magnetopause. This is exactly the behaviour observed by the Pioneer 10 magnetic field experiment⁶.

The Pioneer 10 observations just inside the magnetopause also showed several complete traversals of the current sheet⁶. This is most readily understood if solar wind control of the outermost regions of the jovian magnetosphere was the mechanism responsible, rather than an internal process as required by the alternate models. Thus we conclude that, on the basis of Voyager 1 magnetic field observations, a consistent model of the jovian outer magnetosphere should incorporate

the development of an extended magnetic tail with an embedded neutral sheet.

The observations of the positions of the outbound magnetopause crossings imply a magnetic tail radius of 300–400 R_J , if the tail is assumed to be roughly circular. Figure 4 portrays the distortion of the magnetic field in the current sheet region in the inner magnetosphere and the merging of the magnetodisk current sheet with the neutral sheet of the magnetic tail in the outer magnetosphere. Figure 4 is based primarily on an analogy with terrestrial magnetosphere studies and is adapted for Jupiter by showing a much larger distension of the magnetic field lines near the current sheet. Jupiter also presents a smaller angle of attack for the solar wind, relative to the magnetic equatorial plane of

Table 1 Predicted system III (1965) coordinates of jovian auroral zone					
Lat	Northern Long	B	Lat	Southern Long	B
+81.2°	+143.4°	10.8	-77.3°	8.4°	9.1
+80.6°	+180.4°	10.3	-78.0°	336.7°	9.8
+73.9°	+190.2°	10.6	-80.5°	306.4°	10.2
+67.0°	+184.1°	11.5	-84.2°	261.5°	10.2
+62.7°	+174.2°	12.5	-84.5°	177.7°	10.0
+61.9°	+163.0°	13.2	-79.9°	128.9°	9.6
+64.4°	+151.3°	13.3	-76.1°	102.6°	9.1
+69.5°	+140.0°	12.8	-75.3°	78.1°	8.6
+75.7°	+132.8°	11.8	-76.6°	45.8°	8.6

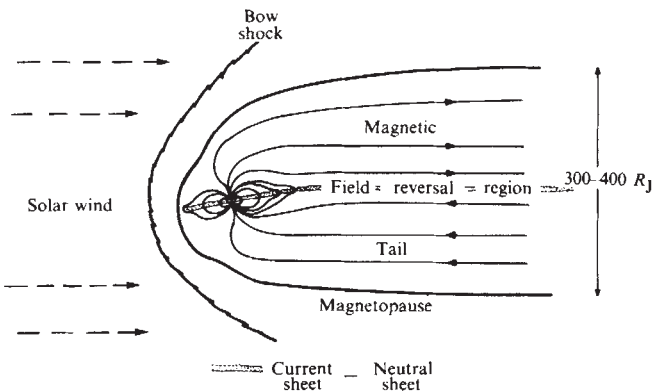


Fig. 4 Sketch of noon-midnight magnetic meridian plane (the $x-Z_{SM}$ plane) geometry of jovian field lines and current sheet-neutral sheet.

Jupiter, because of the small obliquity to the ecliptic of the jovian rotation axis ($\sim 3^\circ$). Nonetheless, the observed curvature of the neutral sheet and current sheet observed by Voyager 1 in the outbound traversal imply a three dimensional geometry so that the tail is not symmetric about the neutral sheet. Voyager 2, which will pass further down the magnetic tail, is expected to provide additional insight on the structure of the jovian magnetosphere, especially the outer portion.

As on Earth, it can be assumed that the magnetic field in the magnetic tail connects to the polar cap regions in order to estimate the size of the polar cap region itself. Figure 5 compares tail field observations near the magnetopause by Voyager 1 and Pioneer 10 with theoretical models which assume a polar cap radius of 8° , 10° and 13° . For this computation, it is assumed that only the planetary magnetic dipole term is responsible for the total flux connecting to the magnetic tail. As seen in Fig. 5, Voyager 1 and Pioneer 10 suggests a polar cap auroral zone of $10^\circ \pm 1^\circ$. This is substantially smaller than Earth's, which is approximately 18° to 22° . This is also much smaller than the polar cap region inferred at Mercury by the same analysis, approximately 18° – 26° .

The Earth's magnetic tail field is known to decrease significantly in magnitude down the tail and to vary in time due to changing solar wind and internal magnetospheric conditions. Thus, due to probable merging across the neutral sheet and flaring of the magnetopause boundaries, the field intensity in the jovian tail is expected to decrease significantly tailward and indeed, the Voyager 1 (3 nT) and Pioneer 10 (5 nT) observations show this to be true.

The implication of the existence of a magnetic tail, with respect to the polar cap region, can be further studied by determining the location of the polar cap auroral zones more precisely. While this cannot be done for an exact model of the jovian magnetosphere, a reasonable approximation can be made. The equivalent dipole co-latitude has been computed by equating the polar cap flux with the total flux in the magnetic tail. With this value, we trace the field lines from the equator at radial distances corresponding to 8° and 13° using the GSFC O_4

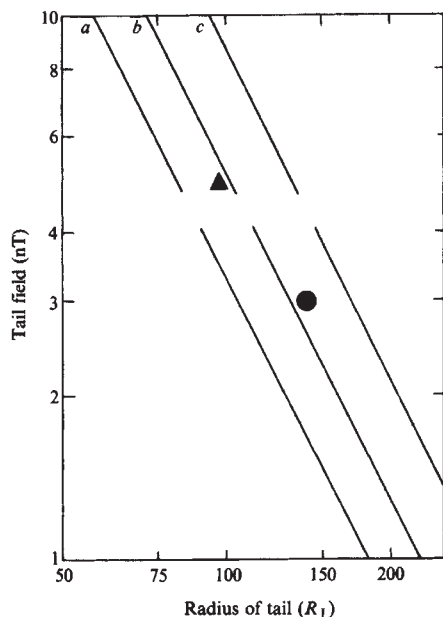


Fig. 5 Relationship between radius of jovian magnetic tail and field intensity, assuming conservation of polar cap magnetic flux in magnetic tail. $\blacktriangle$ Pioneer 10; $\bullet$ Voyager 1. $B_T = 4B_0(R_1/R_T)^2 \sin^2 \theta_{pc}$ where θ_{pc} is the co-latitude: a, $\theta_{pc} = 8^\circ$; b, $\theta_{pc} = 10^\circ$; c, $\theta_{pc} = 13^\circ$.

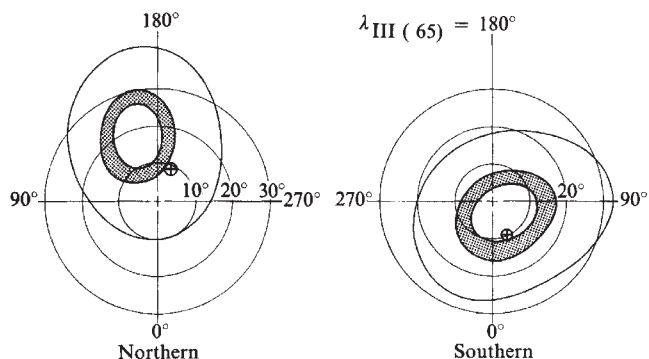


Fig. 6 Jupiter's polar cap auroral zones predicted from the observed magnetic tail and the GSFC O_4 planetary field model. $\oplus$ Magnetic dipole axis; stippled area indicates auroral zone; solid line indicate Io L shell footprint.

planetary magnetic field representation, including quadrupole and octupole terms, but neglecting external currents. The auroral zones so derived are shown in Fig. 6.

The small size of the zones is immediately evident as is a surprisingly large eccentricity of the northern polar cap. Also shown are the footprints of the Io fluxtube, which map out the position of the jovian line of force which threads the satellite Io as it orbits the planet. The asymmetric location of the oval is due to the multipole characteristics of the planetary field while its size is determined by the number of field lines required to connect to the total flux in the magnetic tail. The uniquely eccentric position of the northern polar cap auroral zone of Jupiter is an important and anomalous feature of the jovian magnetospheric configuration. Table 1 presents the coordinates of the polar cap auroral zones, assumed dipole co-latitude of 10° , and field magnitude (in gauss).

These observations show convincing evidence of the development of an extended magnetic tail and imbedded neutral sheet in the nightside magnetosphere of Jupiter, which is formed by the solar wind interaction. This indicates substantial external control of the outer magnetosphere of Jupiter rather than only inner control as postulated by the magnetodisk and magnetic anomaly models. The size of the tail inferred (300–400 R_J diameter), and the field intensity measured (3 nT) imply rather smaller polar cap regions at Jupiter than at Earth. The complex planetary magnetic field model GSFC O_4 yields a highly eccentric northern polar cap region, encircling neither rotation axis nor magnetic axis.

We thank our Voyager colleagues, especially C. K. Goertz for helpful discussions.

Received 27 June; accepted 11 July 1979.

1. Van Allen, J. A. *et al.* *Science* **188**, 459–462 (1974).
2. Smith, E. J. *et al.* *J. geophys. Res.* **79**, 3501–3513, 1974.
3. Goertz, C. K. *Space Sci. Rev.* **23**, 319–344 (1979).
4. Northrop, T. G., Goertz, C. K. & Thomsen, M. F. *J. geophys. Res.* **79**, 3579–3583 (1974).
5. Kivelson, M. G., Coleman, P. J., Froidevanx, L. & Rosenberg, R. L. *J. geophys. Res.* **83**, 4823–4841, 1978.
6. Smith, E. J., Davis, L. Jr & Jones, D. E. *Jupiter* (ed. Gehrels, T. 788–829 (University of Arizona Press, 1976).
7. Dessler, A. J. & Vasyliunas, V. *Geophys. Res. Lett.* **6**, 37–40 (1979).
8. Acuna, M. & Ness, N. F. *J. geophys. Res.* **81**, 2917–2922 (1976).
9. Ness, N. F. *et al.* *Science* **204**, 982–987 (1979).
10. Ness, N. F. *J. geophys. Res.* **70**, 2989–3005 (1965).
11. Ness, N. F., Behannon, K. W., Lepping, R. P. & Whang, Y. C. *J. geophys. Res.* **80**, 2708–2716, 1975.
12. Carr, T. D. & Desch, M. D. in *Jupiter* (ed. Gehrels, T.) 693–737 (University of Arizona Press, 1976).

Departure from rigid co-rotation of plasma in Jupiter's dayside magnetosphere

R. L. McNutt, Jr, J. W. Belcher, J. D. Sullivan, F. Bagenal & H. S. Bridge

Center for Space Research, Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

The MIT plasma experiment on Voyager 1 shows that the plasma flow is not in strict co-rotation at distances greater than 10 jovian radii.

CO-ROTATION of the jovian magnetospheric plasma with the planet to some radial distance has been the basic assumption in all analyses of plasma dynamics in near Jupiter space¹. Brice and Ioannidis² hypothesised that co-rotation should dominate convective motion in the entire magnetosphere. The MIT plasma experiment on the Voyager 1 spacecraft has made the first detailed *in situ* measurements of the low energy (10 eV to 5.95 keV) component of the jovian magnetospheric plasma³. The preliminary analysis of these measurements presented here shows departure of the plasma flow from strict co-rotation at radial distances greater than ≈ 10 jovian radii (R_J). Data used in this analysis were taken in the dayside jovian magnetosphere before the closest approach of the Voyager 1 spacecraft to the planet.

The experimental package consists of four modulated-grid Faraday cups⁴. Three of these form the main sensor cluster and the fourth, or side sensor, is mounted perpendicular to the symmetry axis of this cluster. During that part of the inbound leg of the Voyager 1 encounter which concerns us here, the cluster symmetry axis looks away from Jupiter and towards the Earth while the side sensor looks roughly into the direction of co-rotating plasma flow. In the positive ion mode, the experiment measures currents in both a low resolution mode (the L mode) with $\sim 29\%$ resolution in energy per charge and a high resolution mode (the M mode) with $\sim 3.6\%$ resolution in energy per charge. One L mode spectrum is obtained every 96 s and one full M mode spectrum every 192 s.

We have analysed low resolution mode spectra inbound from 12.00 UT, 4 March, 1979 ($22.6 R_J$) to 02.00 UT, 5 March, 1979 ($11.4 R_J$). By virtue of its wider energy-per-charge channels, the L mode has a higher signal-to-noise ratio than the M mode and gives better quality measurements for the protons in the data set considered. The L mode measurements in the side sensor frequently show good separation between the proton peak and that of the heavy ion constituents although the latter are, in general, unresolved (see Fig. 3 of ref. 3). As long as the plasma is

cold (greater than mach 2) and flows directly into the side sensor, the response function⁵ of the side sensor can be taken as unity to about 5% accuracy. Under this assumption, we have fit two convected, isotropic maxwellians to the side sensor L mode spectra, one to the protons and one to the heavy ions. In general, this provides a good representation of the data. The fit is poor if the mach number becomes so low that the two peaks strongly overlap or sufficiently high that the multi-component nature of the heavy ions begins to appear. Such fits are not presented here.

A more realistic response function for the side sensor is (S. Olbert, personal communication)

$$\exp(-\alpha u_i^2/u_p^2)$$

where α is a constant of order unity, u_i is the ion velocity transverse to the cup normal and u_p is the ion velocity parallel to the cup normal. It can be shown by convolving this response function with a convected, isotropic maxwellian that maxwellian fits assuming unit response always overestimate the plasma bulk speed into the side sensor. The error increases with decreasing mach number of the flow and to a lesser extent with increasing component of plasma bulk velocity transverse to the cup normal. Thus, the speeds we derive from such fits should be taken only as upper limits on the component of plasma bulk velocity into the side sensor.

In Fig. 1, the solid line is the predicted component of rigid co-rotation velocity into the side sensor. The time period covered is that mentioned above with the radial distance of the spacecraft from Jupiter shown along the abscissa. We also plot the component of bulk velocity of the protons into the side sensor, as derived from the analysis described above. Gaps in the data trace appear where the fit is poor. The three circles represent proton measurements from three individual M mode spectra. During this time period, these M mode spectra are somewhat anomalous in that the protons exhibited a strong M mode signal. The spectra give proton and heavy ion bulk speeds consistent with the more numerous L mode measurements; they also show conclusively that the observed departure from co-rotation is not a spacecraft charging effect⁶.

We stress that the plotted quantity is an upper limit. The present analysis is highly dependent on mach number which varies for these spectra between 1 and 4. Hence, structure in the curve should not be overinterpreted. The derived upper limits on the proton bulk speed into the side sensor lie consistently below the value expected for rigid co-rotation in the range from 11 to 22 R_J . We intend to do a more complete analysis of this data set, incorporating data from the main sensor and using the detailed response functions of all sensors. Construction of the full vector velocity of the plasma may then be possible for this magnetospheric region.

We thank S. Olbert and A. J. Lazarus for useful comments. This work was supported in part by NASA contract 953733.

Received 27 June; accepted 11 July 1979.

1. Kennel, C. F. & Coroniti, F. V. *Astr. Astrophys.* **15**, 389–436 (1977).
2. Brice, N. M. & Ioannidis, G. A. *Icarus* **13**, 173–183 (1970).
3. Bridge, H. S. *et al. Science* **204**, 987–991 (1979).
4. Bridge, H. S. *et al. Space Sci. Rev.* **21**, 259–287 (1977).
5. Vasyliunas, V. M. in *Plasma Physics* (Lovbergs, R. H. ed.) Vol. 9B, 49–88 (Academic, New York, 1971).
6. Sullivan, J. D. & Bagenal, F. *Nature* **280**, 798–799 (1979).

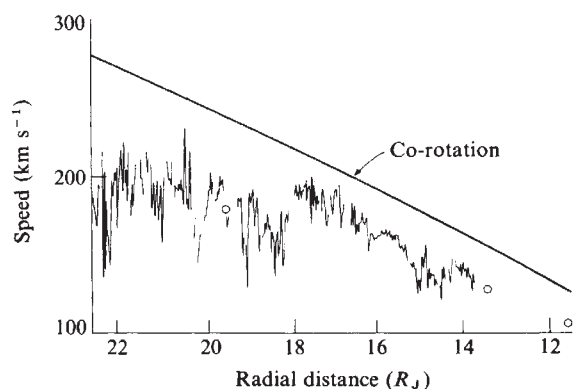


Fig. 1 Upper limit on the component of bulk plasma velocity into the side sensor versus radial distance.

Voyager 1 photopolarimeter experiment and the phase curve and surface microstructure of Ganymede

Kevin Pang*, Charles W. Hord†, Robert A. West†, Karen E. Simmons†, David L. Coffeen‡, Jay T. Bergstralh§ & Arthur L. Lane§

* Planetary Science Institute, 283 South Lake Avenue, Suite 218 Pasadena, California 91101

† Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, Colorado 80309

‡ Goddard Institute for Space Studies, 2880 Broadway, New York, New York 10025

§ Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, California 91103

Optical wavelength photometry used to determine the surface microstructure of Ganymede, shows that its surface is more porous than Callisto.

GANYMEDE, Jupiter's largest satellite, has a density of 1.9 g cm^{-3} , suggesting a bulk composition of about 50% water. Earth-based near-IR spectroscopic studies indicate that a half to two-thirds of its surface is covered with water ice¹. Images of the satellite at a resolution of 2.5 km obtained by Voyager 1 cameras show an astonishing topography of cratered areas and grooved areas². Most of the grooved terrain is a mosaic of systems of sinuous grooves and ridges. Radar echoes of Ganymede show a surprising reversal of the sense of reflected circular polarisation³, indicating that the echoes arise principally from double reflections, and that the surface of Ganymede is therefore quite rough at the wavelength scale (13 cm). Optical wavelength photopolarimetry can reveal the surface characteristics of an atmosphereless object at a still finer scale, that is, its surface microstructure.

A detailed description of the Voyager photopolarimetry experiment has been published elsewhere⁴. The instrument was used as a single-wavelength photometer for whole-disk observations of Ganymede, made over a 3-day period, beginning 5.5 days before the closest approach to Jupiter. Figure 1 shows the brightness of Ganymede at 590 nm as a function of the range to the satellite. Each dot represents the average of 160, 320, 400, 320 and 160 individual raw data measurements, respectively, while the error bar denotes the standard deviation. The background count was essentially zero at such large distances from the radiation belts. As the range to the satellite decreased, its observed brightness did not increase with the inverse square of the distance (the smooth curve in Fig. 1). The instrument was seeing a smaller and smaller lit surface as the Sun-Ganymede-spacecraft angle (solar phase angle) was increasing. Ground-based photometry and Voyager 1 imaging have shown that the surface albedo of Ganymede is inhomogeneous, so that its integrated disk brightness varies with longitude.

It is not possible to separate the phase effect and the rotational effect using the Voyager 1 data alone. However, the rotational brightness variation of Ganymede has been observed from Earth through *UBV* bandpass filters⁵. Fortunately the spectral reflectance of Ganymede is rather flat in the wavelength interval common to the *V*-band (490–650 nm) and the bandpass of the Voyager photopolarimeter (570–610 nm) (see Fig. 2). Consequently we may correct for the rotational effect in our data (covering the longitude range 109°–264°) using published light

curves, that is, Ganymede's nominal visual magnitude versus the rotational angle⁵. The magnitude of Ganymede as observed by the Voyager 1 photopolarimeter, corrected for rotational effect, and ground-based photometric observations⁶ are shown in Fig. 3. These *V*-magnitudes have also been corrected for rotational effect. The data are limited to longitudes between 90° and 270° (the hemisphere facing away from Jupiter). The Voyager data have not been transformed into the *UBV* system but simply normalised to the ground-based observations.

With the extension to 33° phase angle, the shape of Ganymede's phase curve becomes more apparent. Satellite and asteroid phase curves are usually characterised by two parameters: the phase coefficient and the opposition effect. The slope of the straight line of best fit through data points between 6° and the maximum phase angle observable is called the phase coefficient. The magnitude actually observed at zero phase minus that extrapolated from the regression line just defined is called the opposition effect. Ganymede's phase coefficient and opposition effect derived from the data in Fig. 3 are $0.011 \pm 0.001 \text{ mag per deg}$ and $0.17 \pm 0.03 \text{ mag}$, respectively. The error bars represent the statistical scatter in the data, and do not include systematic instrumental errors. There has been a long-term decrease of the instrument's sensitivity during the Voyager 1 mission. The decrease over the 3 days of this observational period is no more than 5%, based on consistency of Jupiter data. Correcting for such a decrease would lower the phase coefficient and increase the opposition effect. These changes are in directions that make the scientific conclusions stronger, as shown below. Other possible sources of systematic error were not checked immediately before Jupiter encounter. We suspect nonlinearity in the instrumental response at high count rates. We expect this effect was less at the count rates observed for Ganymede. However, we are encouraged by the fact that the reduced Voyager data in Fig. 3 are consistent with the phase curve expected from ground-based observations even though the raw data counts span a threefold range.

The derived phase coefficient of Ganymede is remarkably small. Asteroids typically have phase coefficients ranging from 0.020 to 0.055 mag per deg (ref. 7), while the Earth's Moon, Mercury and the martian satellites have values between 0.03 and 0.04 mag per deg (ref. 8). Only Europa^{5,6} seems to have a phase coefficient smaller than that for Ganymede. The value for Ganymede approaches that of a Lambert sphere (a diffusely reflecting sphere, with a phase coefficient of 0.006 mag per deg at 20° phase angle).

Theoretical techniques for interpreting an observed phase curve to derive information on the surface of an atmosphereless object have been discussed by Veverka⁹, Johnson and Matson¹⁰, and others. The discussion below is modelled on that given by

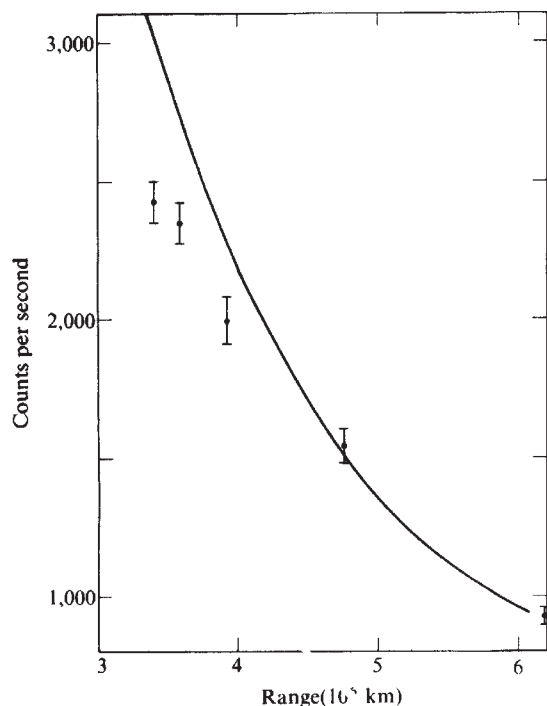


Fig. 1 Whole-disk brightness of Ganymede at 590 nm as a function of the distance between the satellite and the Voyager 1 photopolarimeter. Points are averages of many individual measurements (see text). Error bars are standard deviations. The smooth curve represents brightness variation according to the inverse square law.

Veeverka. In essence there are three scales of surface roughness that determine the shape of an object's phase curve: the macroscopic or large-scale topography, the particulate or crystalline nature (microstructure), and small-scale texture intermediate to these two extremes. Shadowing by topographical features is of major importance at all but the smallest phase angles. The greater the large-scale roughness, the greater will be the phase coefficient. The albedo and scattering phase function of the individual particles or crystals are determined by their size, shape, and complex refractive index. Multiple scattering

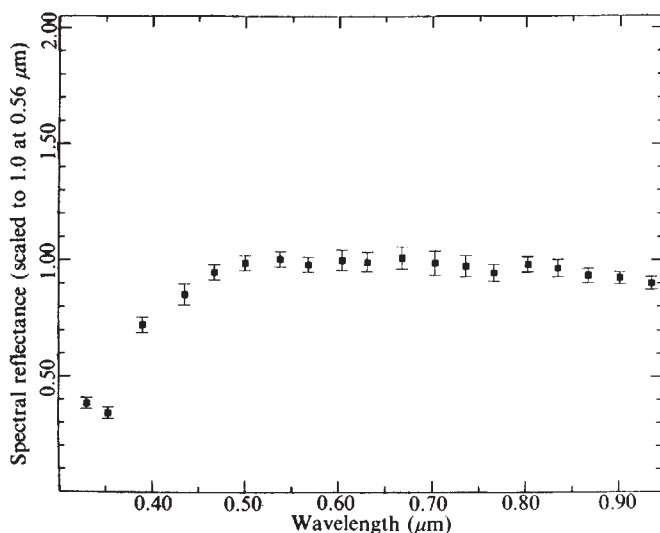


Fig. 2 The relative spectral reflectance of Ganymede as observed from the 224-cm telescope on Mauna Kea Observatory on 28 December 1978. Error bars represent standard deviations of different measurements taken during the same night.

among the particles is generally important, except for the very darkest materials. The greater the particle albedo, the greater the multiple scattering, and the less will be the phase coefficient. For transparent or translucent particles, diffuse transmission and reflection are also involved. The intermediate scale of roughness, the 'texture' of the surface, is determined mostly by the packing characteristics of the particles. The porosity of a surface strongly influences the phase curve at small phase angles.

Modelling of scattering by surfaces is primitive; it is not yet possible to derive surface parameters from observed phase curves only. However, if information on two of the three scales of roughness is available one may set limits on the third. The Hapke-Irvine model for approximating the opposition effect of a dark surface is a case of such an analysis. This model applies to dark particles, for which multiple scattering may be neglected, and to very small phase angles, so that shadowing by large-scale topographic features is not important. With these simplifying assumptions the opposition effect of such a layer is determined

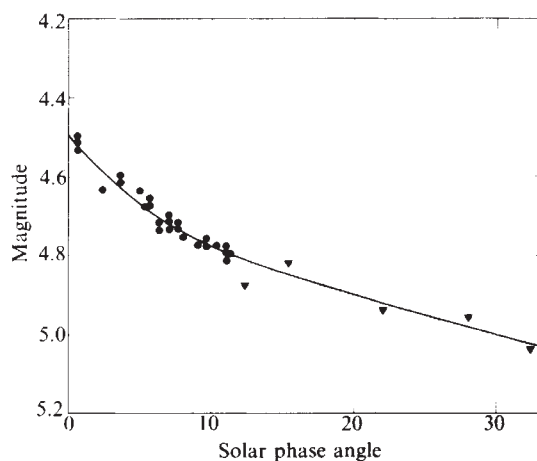


Fig. 3 The phase curve of Ganymede for the hemisphere facing away from Jupiter (longitudes between 90° and 270°). Rotational brightness variations have been corrected for. The Voyager data (▼) are normalised to the ground-based observations (●) of Millis and Thompson⁶.

largely by the packing characteristics, described by the compaction parameter.

$$D = \frac{3}{4\pi} \times \frac{\rho}{\rho_0}$$

where ρ is the density of a macroscopic volume element and ρ_0 is the density of a single particle. The Hapke-Irvine model seems to fit available observations of dark surfaces and seems useful in establishing relative values of D . Veeverka⁹ used this model to compute opposition surges for a variety of compaction parameters. A surface with $D \sim 0.13$ exhibits no opposition effect. As D decreases the surface becomes more porous, and the opposition effect develops. For the leading (trailing) side of Callisto, which has an opposition surge of 0.25 mag (0.13 mag), D would be approximately 0.025 (0.04). The surface material on Callisto thus seems to be rough and porous. All other factors being equal a larger opposition effect means a more porous surface.

Although it is tempting to use the Hapke-Irvine model in analysing the phase curve of Ganymede, we must bear in mind that this model is valid only for dark surfaces like that of the moon (visual albedo 0.12) and Callisto (0.17). The visual albedo of Ganymede, 0.43, is too high for the simplifying assumptions to be valid. However, the topographies of Callisto and Ganymede seem to be similarly rough (Callisto is more heavily cratered but Ganymede is extensively grooved), and the degrees of roughness of the two satellites at 13 cm scale are similar. (In

contrast to Ganymede and Callisto the small-scale texture of Io appears to be smooth. 13-cm roughness estimates are from radar observations by Pettengill.³) The opposition effect of Ganymede (0.17 mag) is slightly larger than that of the trailing side of Callisto (0.13 mag). Now if the topographical and 13-cm scale degrees of roughness of the two satellites are similar and their opposition surges are nearly the same, but Ganymede is 2.5 times brighter than Callisto, then it seems probable that the surface of Ganymede is the more porous of the two. If all three scales of roughness were the same, Ganymede would have a smaller opposition effect because, having brighter surface particles, multiple scattering of light would tend to wash out the effects of shadowing. Increased shadowing effects from a more porous surface on Ganymede compensate for the multiple scattering reduction of the opposition effect.

We arrive at a similar conclusion by comparing the phase coefficients of Ganymede and Callisto. Callisto's phase coefficient (0.025 mag per deg) is more than twice as large as that of Ganymede (0.011 mag per deg). Assuming that the large and intermediate scale roughnesses of the satellites are similar their

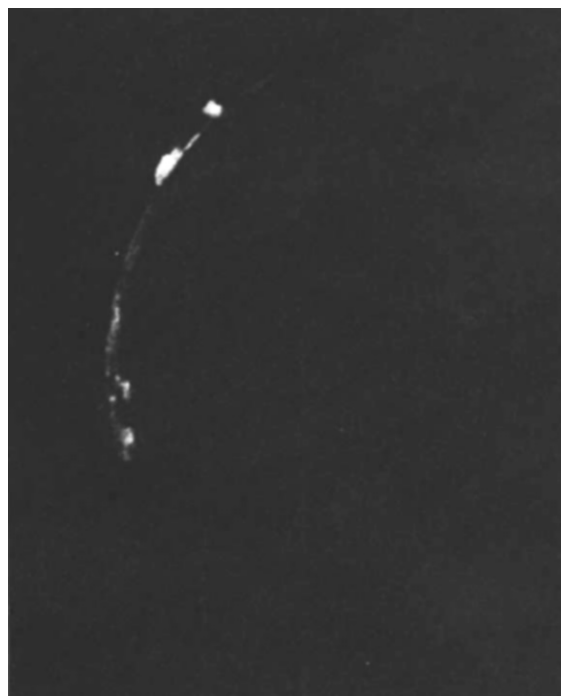
phase coefficients should be the same if their albedos were the same. The greater albedo of Ganymede results from greater single particle albedo, which implies more multiple scattering, therefore a lower phase coefficient. A porous surface layer is necessary to effect the multiple scattering. Ground-based polarisation studies of Ganymede led to a similar conclusion¹¹. Thus, within the limitations of our simplifying assumptions, we conclude that the surface layer of Ganymede (the hemisphere facing away from Jupiter) is porous, and more so than for Callisto (the trailing side). In contrast to Ganymede, the transient south polar cap of Mars has a quasi-specular surface¹². The difference could be due to annual resurfacing on Mars, while micrometeoroids continuously roughen the surface of Ganymede.

We thank members of the Voyager flight team for their support. Ground-based observations were carried out in collaboration with T. B. McCord and staff of the University of Hawaii Institute for Astronomy. Preliminary data reduction was done in the JPL's planetary image facility, under the direction of R. S. Saunders. This research was funded by NASA.

Received 16 July; accepted 31 July 1979.

1. Pollack, J. B., Witteborn, F. C., Erickson, E. F., Strecker, D. W., Baldwin, B. J. & Bunch, T. E. *Icarus* **36**, 271–303 (1978).
2. Smith, B. A. *et al. Science* **204**, 951–972 (1979).
3. Pettengill, G. B. A. *Rev. Astr. Astrophys.* **16**, 265–292 (1978).
4. Lillie, C. F., Hord, C. W., Pang, K., Coffeen, D. L. & Hansen, J. E. *Space Sci. Rev.* **21**, 159–182 (1977).
5. Morrison, D. & Morrison, N. in *Planetary Satellites* (ed. Burns, J. A.) 363–378 (University of Arizona Press, 1977).

6. Millis, R. L. & Thompson, D. T. *Icarus* **26**, 408–419 (1975).
7. Howell, E. *Bull. Am. Astr. Soc.* **8**, 460 (1976).
8. Pang, K. D., Rhoads, J. W., Lane, A. L. & Ajello, J. M. *Nature* (submitted).
9. Veverka, J. in *Planetary Satellites* (ed. Burns, J. A.) 171–209 (University of Arizona Press, 1977).
10. Johnson, T. V. & Matson, D. L. *Bull. Am. Astr. Soc.* **10**, 597 (1978).
11. Veverka, J. in *Planetary Satellites* (ed. Burns, J. A.) 210–231 (University of Arizona Press, 1977).
12. Pang, K. & Hord, C. W. *Icarus* **15**, 443–453 (1971).



VOYAGER I

For copies of the 30 August issue of Nature with 82 pages of results from Voyager I's visit to Jupiter write:

Nature, Macmillan Journals Ltd, Brunel Road, Houndmills, Basingstoke, Hampshire, England RG21 2XS.

Prices including postage:

UK £1.00
 Rest of World US\$3.50 (surface)
 US\$5.00 (air)

Payment may be made in any currency at the prevailing rate of exchange. Cheques should be made payable to Nature.

articles

Voltage-dependent caesium blockade of a cation channel from fragmented sarcoplasmic reticulum

Roberto Coronado & Christopher Miller

Graduate Program in Biophysics and Graduate Department of Biochemistry, Brandeis University, Waltham, Massachusetts 02254

Cs⁺ ions block the K⁺ conductance of a voltage-gated ionic channel derived from fragmented sarcoplasmic reticulum (SR) and incorporated into an artificial phospholipid bilayer. The block follows a single-site titration curve with a voltage-dependent dissociation constant of 18 mM at +50 mV. Blockade observed with the macroscopic conductance can be fully accounted for by the reduction of the apparent single-channel conductance in the presence of Cs⁺. The voltage dependence is explained by assuming that Cs⁺ binds to the open state of a single channel on a site located about 40% of the way through the membrane from the cis side (the side to which the SR vesicles are added). The blocking site is accessible only from the cis side.

ELECTROPHYSIOLOGICAL measurements in a variety of excitable membranes have demonstrated that Cs⁺ ions block K⁺-conducting ionic channels¹⁻⁷. In the case of the well characterised squid axon K⁺ channel, Cs⁺ blockade seems to be due to the size of the ion and the geometry of the channel's inner 'mouth' (refs 4, 8). The blocking mechanism, as outlined by Armstrong⁸, involves the formation of a Cs⁺-site complex inside the channel which prevents the passage of K⁺ ions.

We are attempting the characterisation of a voltage-gated, cation-selective channel present in the sarcoplasmic reticulum (SR) membrane of mammalian skeletal muscle⁹⁻¹². The channel is studied by incorporating SR membrane vesicles into an artificial phospholipid bilayer by a process similar to membrane fusion^{9,10}. Thus, we may conveniently investigate the conductance properties conferred on the planar bilayer by the SR vesicles. Channels are incorporated irreversibly and with at least 98% orientation^{10,11}, allowing the definition of *cis* and *trans* sides of the membrane (the former being defined as the side on which the SR vesicles are added). By using agents that modify the system's properties from only one side, we found that the channel contains a Ca²⁺-blocking site on the *cis* side¹⁰, a transition metal cation binding site on the *trans* side¹¹, sulphhydryl groups accessible from either side¹¹ and a *trans*-facing arginine or lysine residue accessible to attack by pronase-derived alkaline proteinase B (ref. 12).

A membrane containing a large number of SR channels displays a voltage-dependent K⁺ conductance^{10,11}. The steady-state conductance increases *e*-fold per 23 mV (ref. 12) (*trans* side defined as zero voltage). Each individual channel, on the other hand, operates by a mechanism involving a single conducting state in which the probabilities of being in the conducting (open) or non-conducting (closed) states are voltage

dependent; the open-state conductance is independent of voltage and has a value of 130 pS in 0.1 M K⁺ (refs 10, 11). The channels are ideally selective for K⁺ over Cl⁻ and the single-channel K⁺ conductance is 3.5 times the Na⁺ conductance and 26 times the Li⁺ conductance (unpublished). Cs⁺ is unique among the group IA cations in exerting a profound inhibitory effect on the channel's K⁺ conductance. In this article, we discuss this latter observation and propose a mechanism for the inhibition.

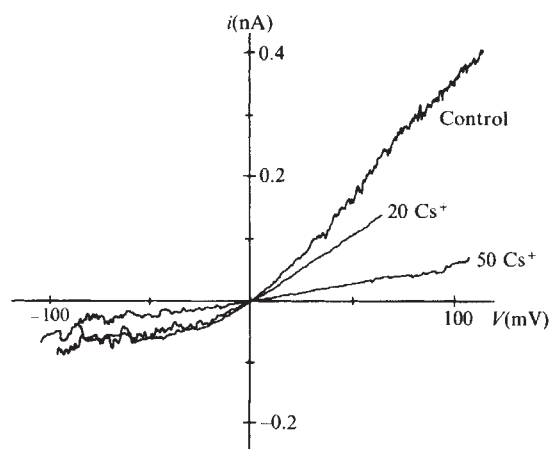


Fig. 1 Steady-state current-voltage relationship and Cs⁺ blockade of the K⁺ current. Planar bilayer membranes were painted on a 1-mm hole, and electrical measurements were made in voltage-clamp conditions^{10,11}; the unmodified membrane conductance was $1-5 \times 10^{-8} \text{ S cm}^{-2}$. The phospholipid mixture was composed of a 50 mg ml⁻¹ decane solution of soybean asolectin (Sigma) and egg phosphatidylethanolamine, in a mol P ratio 1:1. Incorporation of rabbit back muscle SR was accomplished by adding SR protein (50 $\mu\text{g ml}^{-1}$) to the *cis* chamber in 'fusion conditions' (ref. 10). All experiments were done at room temperature (21–23 °C). The aqueous phase was composed of symmetrical solutions of 100 mM K⁺ + *x* mM Cs⁺ + (50 – *x*) mM Li⁺ as the sulphate salts, buffered with 5 mM HEPES-Tris–10 μM EDTA, pH 7.5. The Cs⁺ concentration is indicated in each trace, while the control trace contains no Cs⁺. The current-voltage curves were made in different membranes, and the recordings presented were chosen to have similar conductance at –80 mV (background conductance). They were obtained by applying a slow voltage ramp (20 mV min⁻¹) and recording the current response. Li⁺ was chosen as an 'inert' ion on the basis of its low conductance through the open state of the SR channel and its lack of blocking properties (unpublished). Similar results were obtained replacing Li⁺ by Tris⁺. The Cs⁺ channel conductance is undetectably low (<3 pS in 0.1 M Cs⁺).

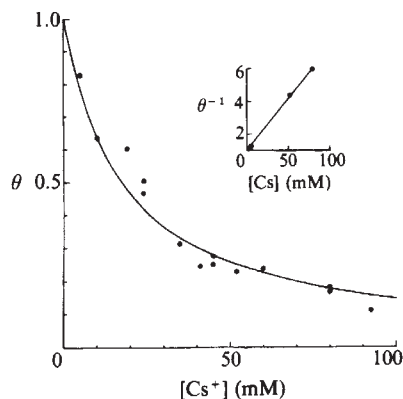


Fig. 2 Titration of Cs^+ block at constant ionic strength. Conditions were as in Fig. 1, except that the aqueous phase was 100 mM K^+ -glucuronate–100 mM Li^+ -glucuronate, 0.1 mM EDTA–10 mM 2[*N*-morpholino]ethane sulfonic acid (MES)-Tris pH 6.7; the lipid mixture was composed of asolectin, 100 mg ml^{-1} in *n*-decane. Cs^+ was added by perfusing the *cis* chamber¹¹ with known volumes of 100 mM K^+ -glucuronate–100 mM Cs^+ -glucuronate, 0.1 mM EDTA–10 mM MES-Tris, pH 6.7. The steady-state conductance at +50 mV and the background conductance were measured before Cs^+ addition. Cs^+ solution was then perfused into the *cis* chamber to the desired Cs^+ concentration. The new conductance values were established not later than 10 s after Cs^+ addition; the background conductance remained unchanged in all experiments, but the +50 mV conductance was reduced after Cs^+ addition. The channel-mediated fraction of conductance remaining after Cs^+ addition, θ , was calculated as

$$\theta(c, V) = [g(c, V) - g_b] / [g(0, V) - g_b]$$

where $g(c, V)$ is the steady-state conductance at a given Cs^+ concentration, c , and applied voltage, V (+50 mV in this case); $g(0, V)$ is the +50 mV conductance before the addition of Cs^+ and g_b is the background conductance and corresponds to the membrane conductance at –50 mV. The solid line corresponds to a single-site binding curve (equation (1)) with $K(V) = 18$ mM. Data are taken from 12 different membranes. Inset: reciprocal plot of equation (1), with data taken from a single membrane. All data were taken from multi-channel membranes with background conductances $> 0.8 \mu\text{S cm}^{-2}$ and +50 mV conductances of the order of $10 \mu\text{S cm}^{-2}$.

Cs^+ inhibition of macroscopic conductance

Although Cs^+ itself does not permeate the channel to any appreciable extent (see Fig. 1 legend), increasing concentrations of Cs^+ eliminate the rectification of the K^+ current (Fig. 1). In the presence of 50 mM Cs^+ , the K^+ conductance at +50 mV is at most 50% higher than the conductance at –80 mV. This latter value serves as a convenient measure of the poorly understood background conductance of the system¹⁰. Control experiments (not shown) demonstrate the complete reversibility of Cs^+ inhibition.

To quantify this inhibition, we studied the effect of various Cs^+ concentrations on the K^+ conductance at a fixed voltage. By maintaining both K^+ concentration and ionic strength constant, we eliminated changes in surface potentials^{13,14}.

Figure 2 shows results of such titrations at +50 mV applied voltage, with Cs^+ added to the *cis* side in place of Li^+ . The inhibition follows a single-site titration curve of the form

$$\theta = \left[1 + \frac{[\text{Cs}^+]}{K_d(V)} \right]^{-1} \quad (1)$$

where θ is the fraction of the channel-mediated conductance not blocked by Cs^+ , and $K_d(V)$ is the macroscopic apparent dissociation constant at that voltage. The pooled data taken from 12 membranes are somewhat scattered around the theoretical curve but the single-site inhibition is clearly seen when data from the same membrane are plotted (Fig. 2, inset). Although *cis* Cs^+

is able to block the conductance at positive and negative voltages, our results indicate that *trans* Cs^+ up to at least 80 mM does not affect the K^+ conductance (data not shown). We will show later that a quantitative estimate for the *cis* and *trans* accessibilities of Cs^+ can be obtained from single-channel fluctuation measurements.

Titration of the single-channel conductance

A probable mechanism for the Cs^+ inhibition of the K^+ conductance of membranes containing many channels is an elimination of the K^+ current through a single channel in the open state, whenever that channel binds Cs^+ . As we could observe single channel fluctuations directly, we tested the above proposal for Cs^+ action by allowing the fusion of only a single SR vesicle with the bilayer^{10–12}. As shown in Fig. 3 inset, we found that Cs^+ produced a gradual reduction in the open-state conductance. The main part of Fig. 3 demonstrates that the open-state K^+ conductance followed a single-site inhibition curve. The apparent dissociation constant measured here is 18 mM at +50 mV.

The inhibitory effect of Cs^+ can be explained on the basis of Woodhull's theory of ionic blockade¹⁵. Following this theory, we shall assume that when Cs^+ binds to a site in the open channel, a non-conducting 'blocked' state is formed. If an equilibrium exists between free and bound Cs^+ , we expect a given channel to fluctuate between three states: closed, open and blocked. We further assume that the lifetime of the blocked state is much shorter than the time resolution of our recording system (~5 ms). Thus, in the presence of Cs^+ , the apparent open-channel conductance corresponds to a time-averaged value determined by the rapid blocking-deblocking rates. It follows that increasing concentrations of Cs^+ should reduce this time-averaged conductance in a single-site binding scheme, as is observed.

The appearance of a short-lived blocked state has been experimentally observed by Neher and Steinbach¹⁶ in the acetylcholine receptor channel of frog muscle when blocked by lidocaine derivatives. An important consequence of the appearance of a blocked state is that the apparent mean open time of the channel (the 'mean group time' of Neher and

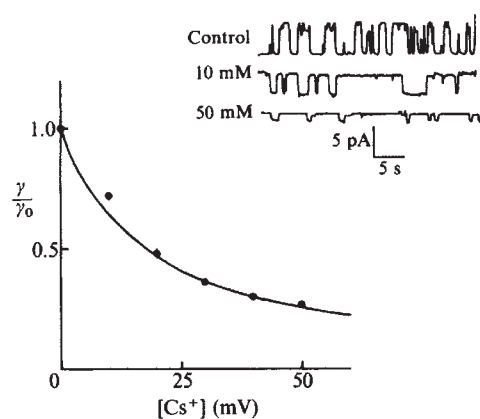


Fig. 3 Cs^+ titration of single-channel conductance. Recordings of single-channel fluctuations were made in conditions described in Fig. 1 at an applied voltage of +50 mV. All the recordings and data points in the graph were made in different membranes. The channel conductance, at a given Cs^+ concentration, was computed from 20–35 conductance fluctuations in 2–4 different membranes. Conductance histograms demonstrate the unitary character of the measured open-state conductance^{10–12}. The solid curve was drawn according to

$$\frac{\gamma}{\gamma_0} = \left[1 + \frac{[\text{Cs}^+]}{K_b(V)} \right]^{-1}$$

where γ is the channel conductance at a given Cs^+ concentration, γ_0 is the channel conductance in the absence of Cs^+ , and $K_b(V)$ is the blocking constant, defined in the text. The curve was fitted with $K_b(V) = 17.8$ mM.

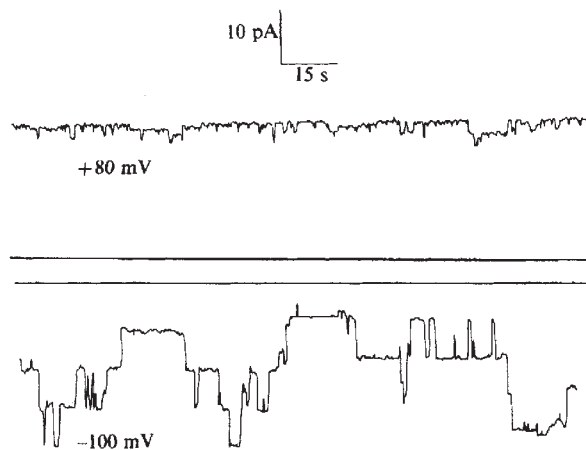


Fig. 4 Single-channel recordings in symmetrical solutions of Cs^+ . Single-channel fluctuations at the indicated voltages were obtained from the same membrane, with 100 mM K^+ buffer and 50 mM Cs^+ in the symmetrical aqueous solutions. The current increases upwards in the +80 mV trace and downwards in the -100 mV trace. The solid lines correspond to the conductance of the unmodified bilayer.

Steinbach¹⁶) should increase with increasing blocker concentrations. Figure 3 (inset) shows that this prediction is fulfilled at least qualitatively; there is a large increase in the mean open times as Cs^+ concentration increases. Although these recordings are from different membranes, they were all taken while only one channel was fluctuating in the membrane.

Cs^+ confers voltage dependence on the single-channel K^+ conductance

In symmetrical solutions containing a fixed Cs^+ concentration, the single-channel K^+ conductance becomes voltage dependent, as indicated in Fig. 4. At 50 mM Cs^+ the channel conductance is 70 pS at -100 mV and only 23 pS at +80 mV. In the absence of Cs^+ , the channel conductance is independent of the applied potential (Fig. 5). Voltage-dependent blockade of ionic channels has been a common finding in electrophysiology^{1-3,15,16}. Although there is no single interpretation for this behaviour¹⁷, the simplest one assumes that the blocking ion moves part of the way across the membrane to form the blocking complex¹⁵. In such conditions, the electrical field acting on the charged ligand would contribute to the standard free energy of the binding reaction¹⁵. Thus, $K_b(V)$, the dissociation constant for the Cs^+ binding reaction, should vary exponentially with the voltage according to

$$K_b(V) = K_b(0) \exp(-\delta FV/RT) \quad (2)$$

where $K_b(0)$ is the zero-voltage dissociation constant, δ is the fraction of the total electrical potential drop across the membrane found at the Cs^+ binding site, and F , R , T have their usual meanings. Combining the single-channel titration curve (Fig. 3) with equation (2), we obtain

$$\frac{\gamma}{\gamma_0} = \left[1 + \frac{[\text{Cs}^+]}{K_b(0)} \exp(\delta FV/RT) \right]^{-1} \quad (3)$$

This expression is similar to that of Adelman and French³ for the Cs^+ blockade of the instantaneous K^+ current in the squid axon K^+ channel, and may be linearised as follows

$$\ln \left(\frac{\gamma_0}{\gamma} - 1 \right) = \ln \frac{[\text{Cs}^+]}{K_b(0)} + \delta FV/RT \quad (4)$$

By measuring γ as a function of voltage, at a constant and symmetrical Cs^+ concentration, we can test whether this model is sufficient to describe the voltage dependence of Cs^+ blockade; in addition, a linearised plot of equation (4) will immediately determine δ and $K_b(0)$. Figure 5 shows the channel conductance measured over a wide range of voltages in the presence of 50 mM Cs^+ (filled circles). Starting from the largest open-state conductance that we were able to measure, about 85 pS at -150 mV, the conductance gradually falls, reaching 15 pS at +150 mV. In the absence of Cs^+ the channel conductance is independent of voltage (open circles), with a value of 92 pS in these conditions. A plot of $\ln[(\gamma_0/\gamma) - 1]$ against voltage is linear with only a slight deviation at positive voltages (see Fig. 5 inset). The parameters obtained from this log plot are: $\delta = 0.35$ and $K_b(0) = 43$ mM. Similar values ($\delta = 0.38$, $K_b(0) = 40$ mM) were obtained from titration experiments as described in Fig. 6. Thus, assuming a constant intramembrane electrical field, we suggest that the site of Cs^+ action is at about 40% of the way across the membrane from the *cis* side.

The observation that in symmetrical solutions of 50 mM Cs^+ , the channel conductance at negative voltages approaches the value for the open channel demonstrates that the blocking site is not readily accessible to *trans* Cs^+ . We have extended the above treatment to the case in which the site is accessible from both

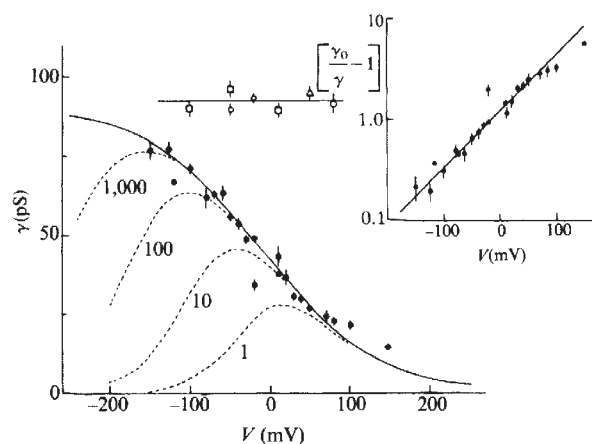


Fig. 5 Voltage-dependent single-channel conductance and *cis-trans* accessibilities of Cs^+ . The figure presents pooled single-channel conductance measurements in symmetrical solutions containing 100 mM K^+ buffer and either 50 mM Cs^+ (filled circles) or other non-blocking cations (open symbols). Over 250 determinations were made, from 15 different membranes. Bars indicate the s.e.m. of the measurement; circles without error bars correspond to the means of less than four determinations. Values for the unblocked channel conductance were determined in 100 mM K^+ buffer plus 50 mM Li^+ ($\circ$), or 25 mM Li^+ + 25 mM Tris^+ (Δ), or 50 mM Na^+ after subtraction of the contribution of the Na^+ conductance (23.6 ± 2 pS ($\pm$ s.e.m.)) ($\square$). Data obtained in 100 mM K^+ buffer plus 50 mM Tris^+ or 50 mM glucosamine⁺ also gave open-channel conductance in the range of 90–95 pS, and are not included. The control channel conductance, from all our Li^+ data, was 92 ± 2 pS ($\pm$ s.e.m.). The solid line in the main graph corresponds to equation (3) in the text, with $K_b(0) = 42.7$ mM and $\delta = 0.35$, values that were obtained from the linearised plot (inset). The family of segmented lines was calculated from a single-binding site model in which *trans* and *cis* Cs^+ reach the site,

$$\frac{\gamma}{\gamma_0} = \left[1 + \frac{[\text{Cs}^+]_{\text{cis}}}{K_b(0)_{\text{cis}}} \exp \frac{\delta FV}{RT} + \frac{[\text{Cs}^+]_{\text{trans}}}{K_b(0)_{\text{trans}}} \exp \frac{(\delta - 1)FV}{RT} \right]^{-1}$$

where $[\text{Cs}^+]_{\text{cis}}$, $[\text{Cs}^+]_{\text{trans}}$ denote the *cis* and *trans* Cs^+ concentrations and $K_b(0)_{\text{cis}}$, $K_b(0)_{\text{trans}}$ are the zero-voltage apparent dissociation constants for *cis* and *trans* Cs^+ binding respectively¹⁵. The curves were calculated with $\delta = 0.35$. $[\text{Cs}^+]_{\text{cis}} = [\text{Cs}^+]_{\text{trans}} = 50$ mM and ratios $K_b(0)_{\text{trans}}/K_b(0)_{\text{cis}}$ are indicated in each curve.

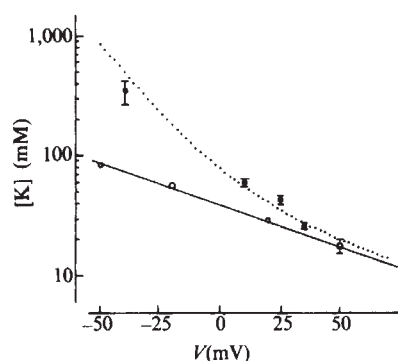


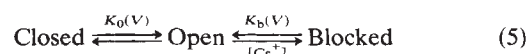
Fig. 6 Comparison of the macroscopic and single-channel blocking constants. All experiments were done in the conditions of Fig. 1. The macroscopic titration (●) was done by Li^+ – Cs^+ exchange as described in Fig. 2; the bars indicate the s.e.m. of the slope from which the macroscopic apparent blocking constant, K_d , is determined. The value of K_d at -42 mV was determined by computing the time-averaged conductance in membranes in which 5–10 channels were fluctuating. The total conductance in this case was obtained by integrating the current trace above the 'zero-level' conductance¹¹, using a digital planimeter. The blocking constant for the macroscopic conductance was obtained, as in Fig. 2, from compiled data from four or more different membranes for each voltage. The single channel dissociation constants were obtained as in Fig. 3. The s.e.m. of the slope from which the channel dissociation constant is determined is contained within the open dots. The value of the macroscopic dissociation constant at $+50$ mV was 18.4 ± 2.7 mM ($\pm$ s.e.m.). The value of the single-channel dissociation constant at that same voltage was 18.2 ± 0.06 mM ($\pm$ s.e.m.). The solid line is the experimentally fitted curve and is drawn according to equation (2), with $K_b(0) = 40$ mM and $\delta = 0.38$. The dotted line corresponds to equation (6) of the text, with $K_b(V)$ as above and $K_0(V) = K_0(0)[\exp(-zFV/RT)]$. The parameters $K_0(0)$ and z , the effective gating charge, were measured as described elsewhere¹²; for these conditions, $K_0(0) = 1.0$, and $z = -1.1$.

sides of the membrane but with different accessibilities, expressed as differences in the apparent dissociation constants for *cis* and *trans* Cs^+ (see Fig. 5 legend). The family of bell-shaped curves shown in Fig. 5 (segmented line) was calculated for different $K_b(0)_{\text{trans}}/K_b(0)_{\text{cis}}$ ratios; thus, it is clear that the Cs^+ binding site is at least 100 times more accessible from the *cis* side than from the *trans* side.

Do the macroscopic and single-channel blockades agree?

When the macroscopic and single-channel dissociation constants are compared in identical experimental conditions (Fig. 6), we observe that they have similar values at positive potentials but are always different at negative potentials. At -42 mV, the macroscopic dissociation constant is about four times larger than the value obtained from single-channel measurements.

The reason for the discrepancy between the two measured 'blocking constants' is apparent from a consideration of the equilibria involved



where $K_b(V)$ is the voltage-dependent dissociation for Cs^+ binding, and $K_o(V)$, defined as the ratio of the number of open to closed channels, is the voltage-dependent equilibrium constant for channel opening¹². The measured single-channel blocking constant reflects only the binding of Cs^+ to the channel, and is equal to $K_b(V)$. In contrast, $K_d(V)$, the macroscopic blocking constant, is determined by both the above equilibria, and it can be shown that

$$K_d(V) = K_b(V) \left(1 + \frac{1}{K_o(V)} \right) \quad (6)$$

Thus, at highly positive voltages, where $K_o(V) \gg 1$, $K_d(V)$ approaches the value of $K_b(V)$; at increasingly negative voltages, where $K_o(V) \ll 1$, the values of the two dissociation constants diverge. The dotted line in Fig. 6 is a theoretical curve for the macroscopic dissociation constant, $K_d(V)$, calculated on the basis of the equilibria described above, in which Cs^+ binds to the site in the open state only. This curve is fully determined by parameters measured independently (see Fig. 6 legend); the agreement between theory and experiment here is thus particularly noteworthy. Based on this departure of the macroscopic dissociation constant from the 'true' single-channel dissociation constant we conclude that the channel must be open before it can be blocked by Cs^+ . In addition, Cs^+ must be released from the site before the channel can close.

Conclusions

We consider that the data presented here are consistent with a blocking model in which *cis* Cs^+ enters the channel, binds to a site and prevents the movement of K^+ through the channel as long as it occupies the site. We have not needed to invoke more complicated phenomena such as current-dependent blocking^{18,19}, multisite blocking^{17,20}, or a Cs^+ binding step before the opening of the channel. Although we have not yet carried out experiments to investigate single-site competition between K^+ and Cs^+ , the chemical similarities between them suggest that Cs^+ may bind to a site normally engaged in the K^+ conduction process.

An important conclusion of this work is that when a blocking reaction is determined by macroscopic steady-state measurements, the observation of a voltage-dependent blocking constant does not imply that the blocking reaction is itself voltage dependent. Only when other voltage-driven equilibria are eliminated, as we have done here by single-channel measurements, can voltage-dependent blocking be demonstrated. In macroscopic electrophysiological measurements, this problem is eliminated by the use of instantaneous, rather than steady-state currents^{3,19}.

Finally, we should emphasise that the SR cation channel is being studied not in its native membrane, but in an artificial phospholipid bilayer. We do not know the extent to which the channel's properties are modified by this reconstitution procedure. Nevertheless, it is gratifying to find in a bilayer system Cs^+ blocking properties so far described only in electrophysiological preparations.

We thank Drs Osvaldo Alvarez and Ramon Latorre for comments and suggestions on this manuscript. R.C. is a Dretzin predoctoral fellow of Brandeis University. This work was supported by NIH grants R01 AM 19826-03 and K04 AM 00354-02.

Received 24 April; accepted 11 July 1979.

- Hagiwara, S., Miyazaki, S. & Rosenthal, N. *J. gen. Physiol.* **67**, 621–638 (1976).
- Gay, L. A. & Stanfield, P. R. *Nature* **267**, 169–170 (1977).
- Adelman, J. Jr & French, R. *J. Physiol., Lond.* **276**, 13–25 (1978).
- Bezanilla, F. & Armstrong, C. M. *J. gen. Physiol.* **60**, 588–608 (1972).
- Hille, B. *J. gen. Physiol.* **61**, 669–686 (1973).
- Lee, K. S., Akaiki, N. & Brown, A. M. *J. gen. Physiol.* **71**, 489–507 (1978).
- Hidalgo, J., Luxoro, M. & Rojas, E. *J. Physiol., Lond.* **288**, 313–330 (1979).
- Armstrong, C. M. in *Lipid Bilayers and Biological Membranes: Dynamic Properties* Vol. 3 (ed. Eisenman, G.) 325–358 (Dekker, New York, 1975).
- Miller, C. & Racker, E. *J. Membrane Biol.* **30**, 283–300 (1976).
- Miller, C. *J. Membrane Biol.* **40**, 1–23 (1978).

- Miller, C. & Rosenberg, R. *Biochemistry* **18**, 1138–1145 (1979).
- Miller, C. & Rosenberg, R. *Biophys. J.* **25**, 25a (1979); *J. gen. Physiol.* (in the press).
- McLaughlin, S. G. A., Szabo, G., Eisenman, G. & Ciani, S. M. *Proc. natn. Acad. Sci. U.S.A.* **67**, 1268–1275 (1970).
- Muller, R. U. & Finkelstein, A. *J. gen. Physiol.* **60**, 285–306 (1972).
- Woodhull, A. M. *J. gen. Physiol.* **61**, 687–708 (1973).
- Neher, E. & Steinbach, J. H. *J. Physiol., Lond.* **277**, 153–176 (1978).
- Hille, B. in *Lipid Bilayers and Biological Membranes: Dynamic Properties* Vol. 3 (ed. Eisenman, G.) 255–323 (Dekker, New York, 1975).
- Adelman, W. J. Jr in *Biophysics and Physiology of Excitable Membranes* (ed. Adelman, W. J. Jr) 274–319 (Van Nostrand Reinhold, New York, 1971).
- French, R. & Adelman, W. J. Jr. *Curr. Topics Membrane Transport* **8**, 161–207 (1976).
- Hille, B. & Schwartz, W. J. *J. gen. Physiol.* **72**, 409–442 (1978).

In vitro formation of a complex between cytoskeletal proteins of the human erythrocyte

E. Ungewickell*, P. M. Bennett, R. Calvert, V. Ohanian & W. B. Gratzer

Medical Research Council Cell Biophysics Unit, King's College, 26-29 Drury Lane, London WC2, UK

The formation of a high-molecular weight complex between spectrin and F-actin depends on the presence of a third cytoskeletal constituent, protein 4.1. Electron microscopy shows that in this ternary complex the actin filaments are linked by bridges, which have the appearance of spectrin. The spectrin must be in the tetrameric state for such bridges to form: the dimer is evidently univalent, for it binds but forms no cross-links. G-actin also fails to form extended complexes. It is inferred that in the native cytoskeleton the spectrin is tetrameric and associated with 4.1 and probably oligomers of actin.

THE unique physical and functional properties which characterise the behaviour of the mammalian erythrocyte are known to be controlled by a cytoskeletal apparatus¹. This is a complex of proteins associated with the cytoplasmic surface of the cell membrane, and its predominant constituent is the high-molecular weight protein, spectrin. However, almost nothing is known of the stoichiometry, patterns of interactions and structural nature of the cytoskeleton. An irregular fibrous network can be discerned in the scanning electron microscope², and a complex of proteins associated with a considerable quantity of lipid, obtained by extraction of erythrocyte ghosts with non-ionic detergent³, probably contains all the elements of the cytoskeleton and evidently retains some of its properties⁴. Another preparation which can be regarded as a fragmented form of the cytoskeleton is the oligomeric complex, containing predominantly spectrin, which is extracted from the membrane into media of low ionic strength⁵. We have observed that both these forms of cytoskeletal complex contain, in addition to spectrin and actin, another membrane protein, termed 4.1 in the indexing system of Fairbanks *et al.*⁶ (referring to mobility in a standard gel-electrophoretic system, containing SDS). This is a protein of apparent molecular weight 78,000, as judged by SDS electrophoresis, and it is present in roughly equivalent proportions to spectrin, taken as a tetramer of molecular weight 960,000. We describe here the requirements for the formation of a complex between the above three constituents of the cytoskeleton, its structural characteristics as seen in the electron microscope and the implications for the structure of the cytoskeleton in the cell.

Formation of the high-molecular weight complex

Earlier work⁷ revealed that preparations of spectrin, chromatographed on Sephadex G-200, and therefore containing dimers and tetramers as well as the oligomeric complex in which the spectrin is associated with actin and 4.1, was able to stimulate the polymerisation of muscle G-actin, and subsequent studies linked this effect to phosphorylation of spectrin⁴. These results could not be reproduced with the dimer or tetramer fraction recovered from Sepharose 4B chromatography, however, and it

became clear that they were a property of the oligomeric fraction (containing endogenous actin and protein 4.1). These results (Pinder *et al.*, unpublished) will be reported in detail elsewhere. They are relevant here only to the extent that they exclude any inference of a binary interaction of high affinity between spectrin and actin. Indeed, our attempts to demonstrate strong binding of spectrin to muscle G- and F-actin by analytical and preparative sedimentation techniques and by column chromatography in a variety of conditions were uniformly unsuccessful, except when the oligomeric complex or a trace of 4.1 was present.

In the present study, the formation of stable high-molecular weight complexes was detected by zone sedimentation in sucrose gradients. Spectrin and protein 4.1 were prepared from fresh human erythrocytes. The spectrin was recovered as the tetramer by extraction in the cold, followed by column chromatography on Sepharose 4B. Protein 4.1 was obtained in pure form from spectrin-depleted ghost vesicles, followed by ion-exchange chromatography on DE-52 cellulose (Fig. 1). Endogenous actin was prepared by extraction of an acetone-treated ghost residue into a medium of low ionic strength⁸ (details are given in Fig. 2 legend). Muscle actin was obtained from rabbit skeletal muscle⁹. These proteins were electrophoretically pure (Fig. 2). When spectrin and F-actin are centrifuged through a sucrose gradient, the protein distribution is indistinguishable

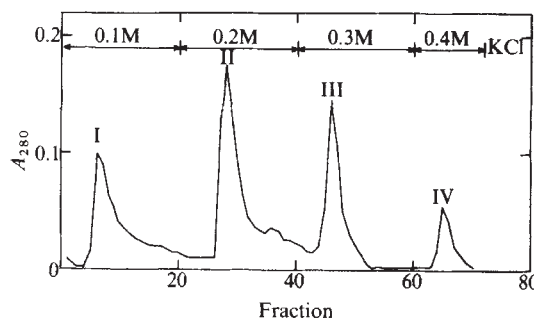


Fig. 1 Preparation of purified protein 4.1 from human erythrocyte membranes. In a typical preparation, 5 ml of packed vesicles remaining after extraction of spectrin at low ionic strength, are further extracted with 0.8 M potassium chloride, 10 mM sodium phosphate, 0.1 mM dithiothreitol and 3 mM phenylmethylsulphonyl fluoride, pH 8.0, in a total volume of 20 ml at 0 °C with stirring for 5 h. The vesicles are then removed by centrifugation at 80,000g for 1 h. The supernatant is dialysed against 7.5 mM sodium phosphate and 0.1 mM dithiothreitol, pH 8.0, and applied to a DE-52 ion-exchange column (10 × 0.9 cm) equilibrated with the same buffer. The column is eluted, first with 40 ml of the initial buffer, followed by two steps with added potassium chloride. A typical elution curve is shown. We have subsequently found that 0.16 M potassium chloride at the second step gives optimal resolution. The haemoglobin and most of band 6 are eluted before addition of salt. The peaks shown in the elution profile comprise: residual band 6 and minor protein constituents (I), protein 4.1 (II), residual spectrin, further traces of 4.1, as well as 2.1 and some minor protein bands (III) and remaining residual spectrin and some 2.1 (IV). The yield of protein 4.1 is 50–60% of the total present in the ghosts.

* Present address: Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138.

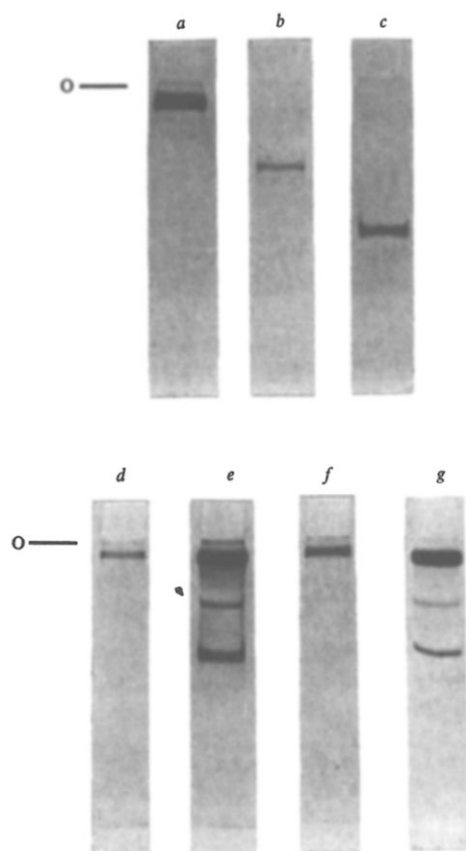


Fig. 2 Demonstration of ternary complex formation: analysis of sucrose gradient sedimentation fractions by gel electrophoresis in the presence of SDS⁶. *a-c* Show the starting materials, which are, respectively, spectrin tetramer, prepared by extraction of ghosts at 4 °C, followed by chromatography on a column of Sepharose 4B as described earlier¹¹, protein 4.1, prepared as described above, and erythrocyte G-actin. The latter is prepared by a method differing in detail from that described by Sheetz *et al.*⁸. In particular, after washing the ghosts from 60–70 ml packed human erythrocytes with 0.5 M potassium chloride and 10 mM Tris, pH 7.4, they were twice more washed with 1 mM Tris, 0.1 mM ATP and 0.1 mM dithiothreitol, pH 7.4, rather than distilled water. The acetone powder⁸ was freed of excess acetone by suction on a medium sinter, rather than centrifugation, and was extracted according to the method of Spudich and Watt⁹. The actin was purified by three successive polymerisation-depolymerisation cycles (1–2 mg yield). Rabbit muscle G-actin was prepared by the method of Spudich and Watt⁹. To form the complex, the proteins were mixed and left for 60 min on ice in 0.1 M sodium chloride, 30 mM Tris, 0.1 mM magnesium chloride and 0.5 mM dithiothreitol, pH 8.0. The protein concentrations in the final mixture were: spectrin 1 mg ml⁻¹, F-actin 0.1 mg ml⁻¹, and protein 4.1 0–0.1 mg ml⁻¹. Aliquots of 0.2 ml of the mixture were applied to 4.5 ml linear sucrose gradients (10–30% w/v) in the above buffer, with a 0.5-ml cushion of 45% sucrose. After centrifugation at 4 °C and 100,000g for 14.5 h in a Beckman SW50 rotor, fractions were collected; pelleted proteins were dissolved in 0.2 ml of 0.1% SDS. Aliquots of 20–50 µl were analysed on the gels, which show the pellet fraction collected from a protein mixture of: *d*, 1 mg ml⁻¹ spectrin and 0.1 mg ml⁻¹ erythrocyte F-actin, with no 4.1; *e*, 1 mg ml⁻¹ spectrin, 0.1 mg ml⁻¹ F-actin and 0.1 mg ml⁻¹ protein 4.1; *f*, the same, with erythrocyte G-actin in place of F-actin; *g*, the same mixture as in *e*, but with muscle in place of erythrocyte F-actin. The origin of migration is indicated.

from the superimposed patterns of the two proteins sedimenting separately. Similarly, there is no evidence from these experiments of the formation of a stable complex between spectrin and protein 4.1 alone. However, when all three proteins are mixed, a rapidly sedimenting component is obtained which, with an increasing proportion of 4.1, ultimately incorporates all the spectrin (Fig. 3). The composition of this stable ternary complex is reflected by the gel electrophoresis patterns in Fig. 2e.

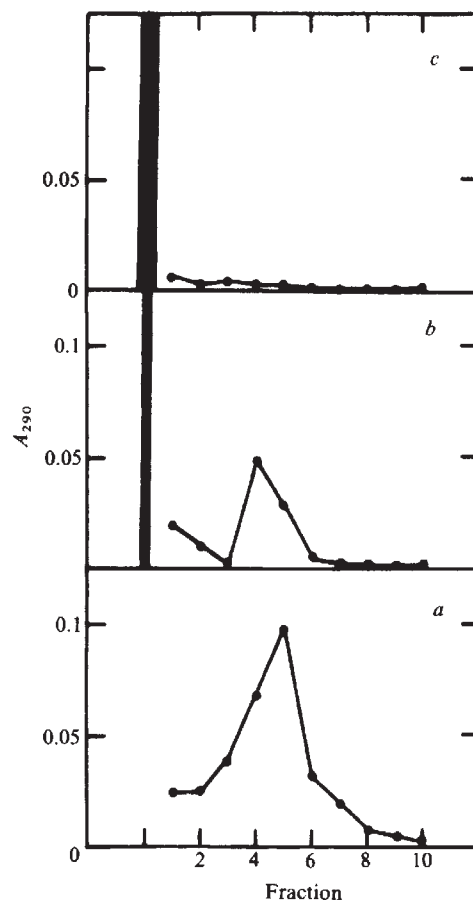


Fig. 3 Sucrose gradient velocity sedimentation of mixtures containing 0.5 mg ml⁻¹ spectrin, 0.1 mg ml⁻¹ erythrocyte F-actin and no added protein 4.1 (*a*), 30 µg ml⁻¹ 4.1 (*b*) or 75 µg ml⁻¹ 4.1 (*c*). The shaded region is the pellet. The experimental details are as described in Fig. 2 legend. Sedimentation is from right to left. In these conditions the presence of 75 µg ml⁻¹ protein 4.1 is sufficient to cause the incorporation of all the spectrin (0.5 mg ml⁻¹) into a high-molecular weight complex.

Complexes containing actin, spectrin tetramer and protein 4.1 in the molar ratio of about 3:1:1 have been obtained, but studies of the possible range of stoichiometries are still in progress. At all events, the results at this ionic strength show the assembly of a complex only if all three cytoskeletal constituents are present. Furthermore, when the actin is kinetically trapped in the depolymerised state (G-actin), at a concentration sufficiently low that no significant polymerisation occurs within the time scale of the experiment, only very small traces of a rapidly sedimenting complex appear (Fig. 2). This does not demonstrate that G-actin is incapable of forming a ternary complex with the other components, but does show that in these (physiological) solvent conditions and in the accessible concentration range, it is not able to participate in the formation of an extended network.

The results shown in Figs 2e and 3 were obtained with purified erythrocyte actin. Muscle actin behaves in a similar manner (Fig. 2g). Muscle actin was also used in the experiment shown in Fig. 4, which demonstrates the progressive incorporation of spectrin and F-actin into the complex with the addition of increasing proportions of 4.1. In Fig. 4a the small amount of complex formed reflects the concentration of residual 4.1, present in the spectrin preparation.

The formation of ternary complexes between cytoskeletal proteins reveals itself in other types of experiment, which also show that the formation of a network requires that spectrin be in the tetrameric state. In a mixture containing F-actin (1 mg ml⁻¹), protein 4.1 (0.05 mg ml⁻¹) and spectrin tetramer (0.4 mg ml⁻¹)

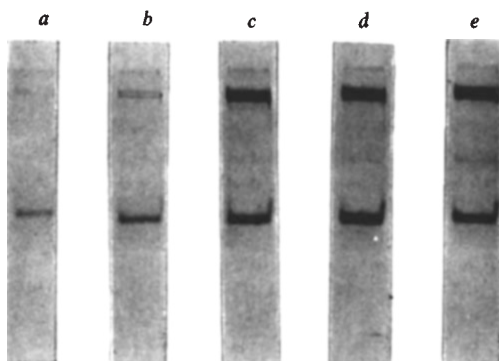


Fig. 4 Gel electrophoresis of the ternary complex, recovered as a pellet after sucrose gradient centrifugation (conditions as in Fig. 2 legend), as a function of the proportion of protein 4.1. In this case muscle F-actin was used. The total spectrin tetramer and F-actin concentrations were kept constant at 1 and 0.1 mg ml⁻¹, respectively. In *a* the pellet contains the complex formed as a result of the residual trace of 4.1 present in the spectrin. In *b-e* purified protein 4.1 was added at a concentration of 16, 40, 80 and 120 µg ml⁻¹, respectively.

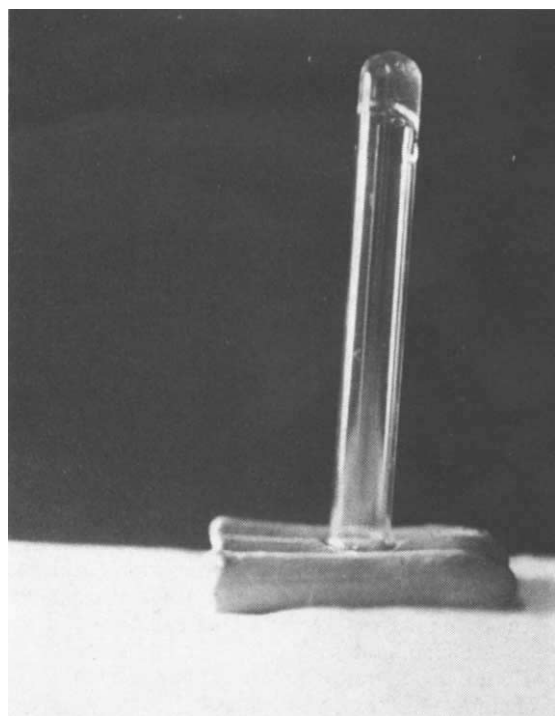


Fig. 5 Gel formed from a mixture containing 0.9 mg ml⁻¹ muscle F-actin, 0.4 mg ml⁻¹ spectrin tetramer and 0.05 mg ml⁻¹ protein 4.1 (all purified as described in Fig. 1 legend). The medium contained 50 mM sodium chloride, 10 mM imidazole, 1 mM ATP, 1 mM dithiothreitol, 1 mM EGTA and 1 mM magnesium chloride, pH 7.2. No gel is formed in the absence of protein 4.1, in the absence of spectrin, or with the dimer form of spectrin in place of the tetramer.

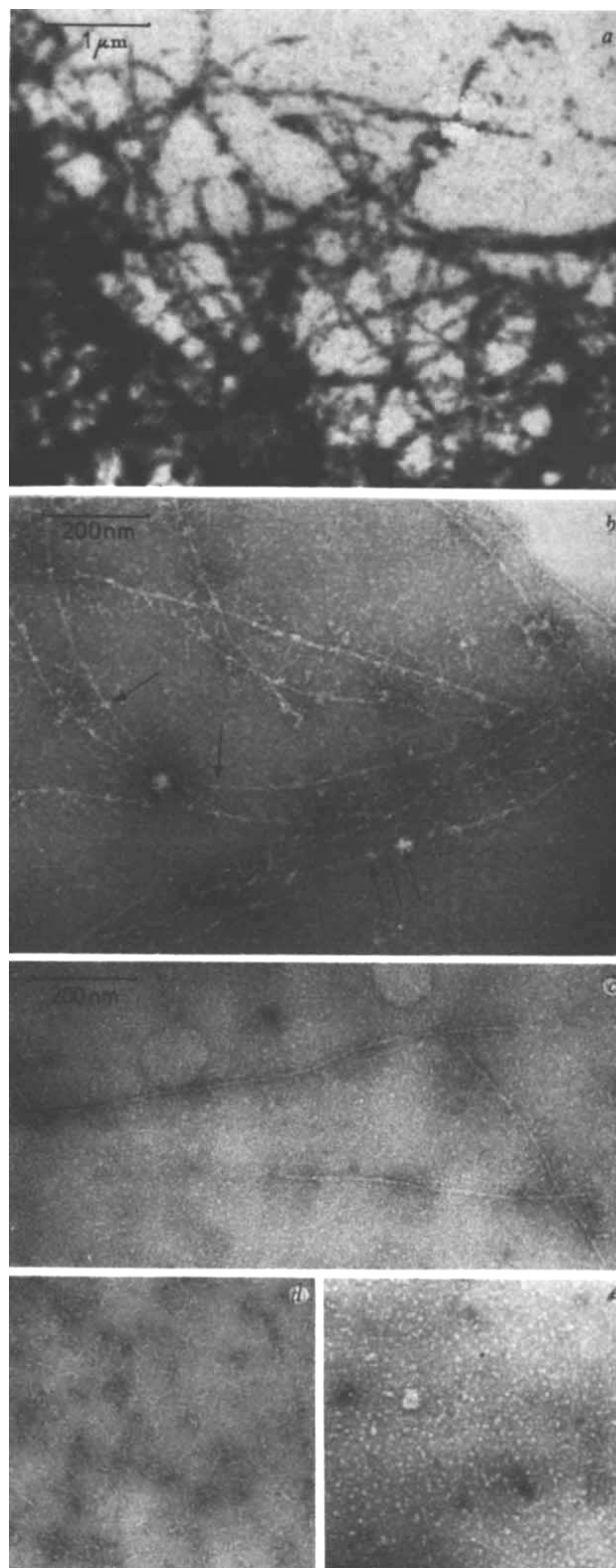


Fig. 6 Appearance in the electron microscope of complexes of F-actin with spectrin tetramer and protein 4.1. Complexes were prepared at a total protein concentration of about 1 mg ml⁻¹. After 10-fold dilution they were applied to carbon-coated grids, negatively stained with 1% uranyl acetate and examined in a Philips EM 200 instrument at 80 kV. *a*, Part of a field of complexed material showing a network of connected fibres. *b*, Complex shown at higher magnification, prepared as in *a* and freed of uncomplexed components by passage through a column (10 × 1 cm) of Sepharose 4B. The actin filaments are associated with thin flexible fibres, which form curved bridges where neighbouring filaments lie close together. Some of these are marked by arrows. *c*, Complex formed with spectrin dimers in place of tetramers in the ternary system. Appendages can be seen on the actin filaments, but no bridges between filaments. *d*, Spectrin tetramers alone, observed in the same conditions. *e*, Preparation of purified protein 4.1.

in 10 mM phosphate, 0.1 M sodium chloride, 2 mM magnesium chloride and 0.1 mM dithiothreitol, a gel is formed (Fig. 5). When two such mixtures, containing respectively spectrin dimer and tetramer, were centrifuged for 15 min at 20,000g to bring down high-molecular weight complex, 15 and 60% of the total protein was recovered as a pellet in the two cases. No gelation occurs in any binary combination of the proteins, nor when the actin is unpolymerised or the spectrin dimeric. The presence of the ternary complex in the cytoskeleton is presumably responsible for the failure of more than a part of the erythrocyte actin to extract into the low ionic-strength buffer in the G-form from the acetone powder. The remainder stays, together with spectrin and 4.1, in the pellet.

Electron microscopic evidence for ternary complex formation

Figure 6 shows the appearance of the complex in the electron microscope. Adjoining F-actin filaments are linked at intervals by curved bridges, leading to a separation in these regions of usually about 50 nm. In some cases the bridges seem to be more extended, corresponding to an inter-filament separation of up to 100 nm. The length and aspect of these bridges are consistent with their identification as spectrin tetramers as they appear when negatively stained (Fig. 6d). The tetramers linking the actin filaments are evidently not as extended as they appear after deposition from solutions containing 70% glycerol and shadowing. Shotton *et al.*¹⁰ have shown that in these conditions the tetramers have a shadowed length of nearly 200 nm. By contrast, protein 4.1 when negatively stained is seen as a compact and presumably globular species (Fig. 6e). The simplest model for the complex, consistent with its appearance in the electron microscope, would involve the attachment of spectrin tetramers, through sites at their extremities, to 4.1 and actin. That the spectrin tetramers should in fact be bifunctional species follows from the electron microscopic observations of Shotton *et al.*¹⁰, for the tetramer is evidently derived from the dimer by end-to-end association, and studies on self-association of the dimer¹¹ have shown that no (isodesmic) associated states higher than the tetramer are formed; in other words that, with the individual polypeptide chains of each dimer apparently lying parallel¹⁰, the self-association sites are internally satisfied (D_2 symmetry).

The postulated structure of the complex seen in Fig. 6 implies that the spectrin must be divalent if a network is to be formed, and that the dimer will not suffice. Parallel experiments with the purified dimer, which is indefinitely stable in the cold¹¹, lead to no such networks as those shown in Fig. 6b, although binding evidently occurs. In the electron microscope the spectrin can be recognised as threads attached to the F-actin filaments (Fig. 6c), but as expected from the present results, no bridges can be discerned.

The interactions leading, as we have described, to the formation of stable ternary complexes of the cytoskeletal proteins can be readily reconciled with several earlier observations, including our own (already referred to) on the efficacy of the oligomeric complex, as opposed to pure spectrin, in inducing polymerisation of G-actin. The formation of a complex between muscle F-actin and (crude) spectrin was, in fact, first inferred by Tilney and Detmers¹² from an increase in viscosity accompanying mixing. Moreover, the attachment of actin filaments to erythrocyte membrane vesicles, only when these contain the oligomeric complex, has recently been reported by Cohen and

Branton¹³. Puszkin *et al.*¹⁴ have described the association of G-actin, evidently by way of polymerisation, to unfractionated spectrin (eluted from Sephadex G-100 and therefore containing the oligomeric complex) adsorbed to latex beads.

Clearly, we cannot rule out the possibility (indeed it must be postulated) that binary interactions between spectrin, actin and 4.1 occur, and these might be expected to reveal themselves in solvents of lower ionic strength. It should be recognised that the removal of all traces of 4.1 from spectrin is extremely difficult, and that even material purified on a Sepharose 4B column contains detectable amounts. (This can probably be ascribed to a self-associated form, the elution profile of which overlaps that of spectrin.) Because few cross-links are required to generate a fraction of high molecular weight, which will manifest itself in hydrodynamically based assays, such spectrin preparations are expected to exhibit artefactual binding to F-actin. This would be most obtrusive when sensitive methods such as low-shear viscometry are used. We consider it likely that a recent brief report of such interactions¹⁵ can be explained in these terms.

Discussion

Our results agree with our earlier conclusion¹¹ that the spectrin is present in the cytoskeletal network *in situ* as a tetramer, rather than a dimer. They also imply that actin in the erythrocyte cytoskeleton is unlikely to be in the monomeric G-form, and that models based on such a view¹⁶ are over-simplifications. However, neither is it likely to be present as long filaments, as in the cytoplasm of many cells, for the quantity of actin present in the erythrocyte is so small that such filaments, even if arranged in a uniform two-dimensional net, would leave such large areas between the intersections that it would be hard to envisage the functional utility of such an arrangement. The inference seems inescapable that the actin is present as very short oligomeric segments. This would also be compatible with the character of the oligomeric spectrin-actin-4.1 complexes extracted from the membrane, which contain no F-actin filaments, and would also explain the failure of electron microscopic examination to detect such filaments in ghosts or in the cytoskeletal complexes prepared by extraction with non-ionic detergent. The only evidence known to us which suggests the presence of actin in the self-associated state, comes from cross-linking experiments¹⁷, in which oligomeric actin species are indeed found. If we assume a 50-nm end-to-end distance for the spectrin tetramers, the amounts of the proteins present in the cell would be sufficient to cover the cytoplasmic surface with (for example) a square lattice, constructed of spectrin tetramers, with connecting elements at the junctions, containing 5–10 actin monomers.

Many other aspects of the protein interactions in the cytoskeleton remain unclear. Thus, whereas there is evidence to link phosphorylation of spectrin to changes in its interaction with its neighbouring proteins^{4,18} and with the control of cell shape¹⁹, we have no evidence that it modifies the formation of the ternary complexes which we have described. Further investigations in this regard are in progress. We note finally that another protein constituent, which migrates in SDS gels in a position very close to that of actin, and therefore probably has a subunit molecular weight of about 45,000, is found associated with the cytoskeleton complex, regardless of how it is prepared. We still know nothing about the nature and function of this species.

E.U. was the recipient of a long-term fellowship from EMBO, and R.C. of a SRC training studentship.

Received 5 April; accepted 19 July 1979.

- Mohandas, N. & Shohet, S. B. *Curr. Topics Haemat.* **1**, 72–150 (1978).
- Hainfeld, J. F. & Steck, T. L. *J. supramolec. Struct.* **6**, 301–311 (1977).
- Yu, J., Fischmann, D. A. & Steck, T. L. *J. supramolec. Struct.* **1**, 233–248 (1973).
- Pinder, J. C., Bray, D. & Gratzel, W. B. *Nature* **270**, 752–754 (1977).
- Lux, S. E. & John, K. M. in *Cell Shape and Surface Architecture* (eds Revel, J. P., Henning, U. & Fox, C. F.) 481–491 (Liss, New York, 1977).
- Fairbanks, G., Steck, J. L. & Wallach, D. F. H. *Biochemistry* **10**, 2606–2617 (1971).
- Pinder, J. C., Bray, D. & Gratzel, W. B. *Nature* **258**, 765–766 (1975).
- Sheetz, M. P., Painter, R. G. & Singer, S. J. *Biochemistry* **15**, 4486–4492 (1976).
- Spudich, J. A. & Watt, S. J. *J. biol. Chem.* **246**, 4866–4871 (1971).

- Shotton, D., Burke, B. & Branton, D. *Biochim. biophys. Acta* **536**, 313–317 (1978).
- Ungewickell, E. & Gratzel, W. B. *Eur. J. Biochem.* **88**, 379–385 (1978).
- Tilney, L. F. & Detmers, L. G. *J. Cell Biol.* **66**, 508–520 (1975).
- Cohen, C. M. & Branton, D. *Nature* **279**, 163–165 (1979).
- Puszkin, S., Maimon, J. & Puszkin, E. *Biochim. biophys. Acta* **513**, 205–220 (1978).
- Brenner, S. L. & Korn, E. D. *Fedn Proc.* **38**, 469 (1979).
- Pinder, J. C., Ungewickell, E., Bray, D. & Gratzel, W. B. *J. supramolec. Struct.* **8**, 439–445 (1978).
- Wang, K. & Richards, F. M. *J. biol. Chem.* **250**, 6622–6626 (1975).
- Wolfe, L. C. & Lux, S. E. *J. biol. Chem.* **253**, 3336–3342 (1978).
- Birchmeier, W. & Singer, S. J. *J. Cell Biol.* **73**, 647–659 (1977).

Nucleotide sequence of the gene coding for the major protein of hepatitis B virus surface antigen

Pablo Valenzuela, Patrick Gray, Margarita Quiroga, Josefina Zaldivar, Howard M. Goodman & William J. Rutter

Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, California 94143

DNA extracted from hepatitis B virus Dane particles has been cloned in bacteria using a plasmid vector. A full-length clone has been examined by restriction endonuclease analysis, and the nucleotide sequence of an 892-base pair fragment from cloned hepatitis B viral DNA encoding the surface antigen gene is reported. The amino acid sequence deduced from the DNA indicates that the surface antigen is a protein consisting of 226 amino acids and with a molecular weight of 25,398. The portion of the gene coding for this protein apparently contains no intervening sequences.

VIRAL HEPATITIS is a major worldwide public health problem. Its aetiology is associated with at least three distinct viral families: hepatitis A, hepatitis B and non-hepatitis A/hepatitis B (ref. 1). The characterisation of the viruses, the pathology associated with the infection and the development of effective means of control have been severely hampered by the very narrow host range specificity of the viruses, and the inability to replicate the virus and obtain cytopathology in tissue culture systems. Nevertheless, progress has been made in identification and preliminary characterisation of the viruses from the sera of infected individuals and from livers at autopsy. The hepatitis A virus is apparently a picornavirus detected as a 27-nm particle^{2,3}. In addition to being infective in humans and chimpanzees, it has recently been transferred to marmosets and has been replicated in tissue culture⁴⁻⁶.

The plasma of individuals infected with hepatitis B show three major particulate structures containing the antigenic determinants apparently specified by the hepatitis viral genome. These include the predominant 22-nm particles, 22-nm filaments of various lengths, and the 42-nm spherical form known as the Dane particle^{7,8}. The Dane particle is probably the hepatitis B virion. Recently, a human hepatoma cell grown in tissue culture was shown to produce small quantities of hepatitis B antigens⁹.

In instances where productive infections and high titres of viruses cannot be obtained in alternative hosts or tissue culture cells, recombinant DNA methods can be used to obtain large quantities of the virus genome for characterisation and biological activity studies. Structural analysis of the various antigens may allow identification and sequence determination of the various genes of the virus. This information, plus the availability of the viral genome, will greatly aid the study of the pathology and eventually the development of antiviral therapy. In particular, it may be possible, using segments of the viral genome, to produce appropriate viral antigens in alternative hosts such as bacteria or yeast for the development of a vaccine.

Hepatitis B is a particularly attractive paradigm for the development of this approach. The hepatitis B viral genome is relatively simple, perhaps consisting of only 3,200 bases^{10,11}. The surface coat, which can be removed by treatment with detergents, contains two main polypeptides, P1 (22–24,000 MW) and P2 (25–29,500 MW)¹². These molecules are antigenically indistinguishable and protein components are probably identical. Their amino acid compositions are the same

within the limits of the analyses used, and the first 19 residues from the N-terminus and the three amino acid residues at the carboxy-terminus seem to be identical¹³. The higher molecular weight molecule (P2) is a glycoprotein, and it has been suggested that the major virus-specific surface antigens are contained within a single gene. Five to seven other polypeptides ranging upward in size to 97,000 MW have been reported to be present in surface antigen preparations¹⁴. It is not known whether any or all of these molecules are coded for by the viral genome. Viral core preparations contain a major polypeptide of MW 17,000 to

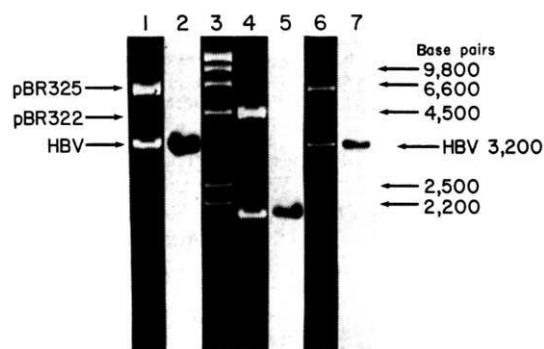


Fig. 1 Cloning and hybridisation of hepatitis B viral DNA. Double-stranded DNA was synthesised in Dane particles (isolated from human sera by Merck Laboratories) by the endogenous DNA polymerase reaction¹⁰. Incubation was for 3 h at 37 °C in a reaction mixture containing Dane particles (250 ng of DNA), 50 mM Tris-HCl (pH 7.5), 0.1 M NaCl, 20 mM MgCl₂, 40 mM NH₄Cl, 0.4% NP40, 10 mM 2-mercaptoethanol and 0.2 mM each dATP, dGTP, dCTP and dTTP. DNA was isolated by phenol extraction and ethanol precipitation after digesting the Dane particles for 60 min at 56 °C with proteinase K (0.8 mg ml⁻¹). Dane DNA was digested to completion with endonuclease *Eco*RI (producing a single 3,200-base pair fragment) and with *Bam*HI (yielding two fragments of 2,100 and 1,100 base pairs). Digested Dane DNA (250 ng) was ligated to the appropriate plasmid (50 ng of *Eco*RI digested pBR325 (ref. 18), or 680 ng of phosphatase-treated, *Bam*HI-digested pBR322 (ref. 30)) for 15 h at 14 °C in a reaction mixture containing 50 mM Tris-HCl (pH 8), 1 mM ATP, 10 mM MgCl₂, 20 mM dithiothreitol and 1 unit of T4 DNA ligase (New England Biolabs). The reaction mixture was used to transform *E. coli* HB101 (in P3/HV1 containment conditions) or *E. coli* λ1776 (in P2/HV2 containment conditions). Recombinant colonies derived from the DNA ligated to the *Bam*HI site of pBR322 were selected by their ampicillin resistance and tetracycline sensitivity¹⁷ and screened for plasmid size¹⁹. A clone (pHBV-2100) of approximately 6,500 base pairs was identified by a modified toothpick assay¹⁹. Recombinant colonies derived from the DNA ligated to the *Eco*RI site of pBR325 (ref. 18) were selected by their sensitivity to chloramphenicol. A clone (pHBV-3200) containing plasmid DNA of approximately 8,600 base pairs was identified by the toothpick assay. Supercoiled plasmid DNA was isolated from either clone by a cleared lysate procedure³¹, purified by CsCl-ethidium bromide gradient centrifugation³¹, digested with *Eco*RI or *Bam*HI and analysed in 1% agarose gels. DNA fragments were hybridised according to Southern³² to Dane particle DNA labelled by nick translation³³. Lane 1, ethidium bromide staining of fragments derived from digestion of pHBV-3200 with *Eco*RI. Lane 2, hybridisation of the same fragments to ³²P-Dane particle DNA and autoradiography. Lane 3, *Hind*III-digested λ DNA as MW standards. Lane 4, ethidium bromide staining of fragments derived from digestion of pHBV-2100 with *Bam*HI. Lane 5, hybridisation of the same fragments to ³²P-Dane particle DNA and autoradiography. Lane 6, ethidium bromide staining of fragments derived from digestion of pHBV-64 (a clone of pHBV-3200 viral DNA inserted in the *Pst*I site of pBR322). Lane 7, hybridisation of the same fragments to ³²P-Dane particle DNA and autoradiography.

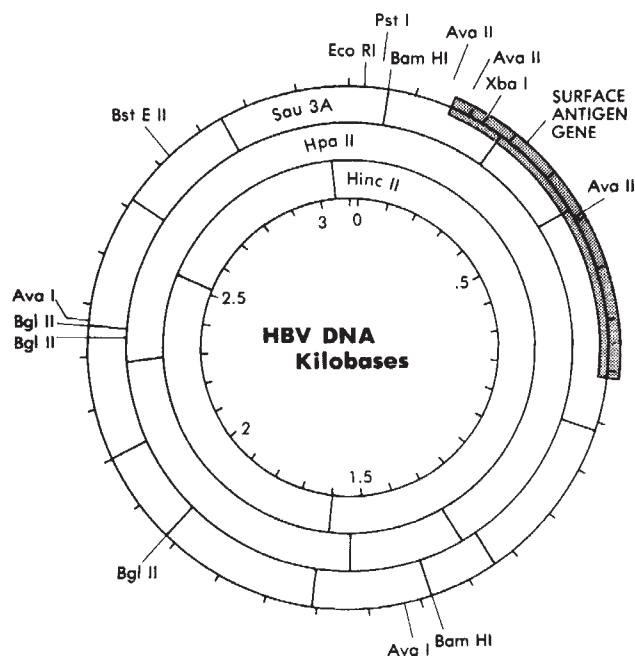


Fig. 2 Restriction endonuclease cleavage map of cloned HBV DNA in plasmid pHBV-3200. Cleavage with restriction endonucleases was carried out in the conditions described by the suppliers with excess amounts of each enzyme. Analysis of the resulting fragments was by electrophoresis in 7% or 10% acrylamide slab gels in 50 mM Tris-borate (pH 8.3) and 1 mM EDTA or in 1–2% agarose slab gels in 50 mM Tris-acetate (pH 8.3) and 1 mM EDTA.

19,000 and a variety of other larger polypeptides of MWs 25,000 to 200,000 (ref. 12). Whether these polypeptides are virus-specific determinants is unknown. A third class of antigens associated with hepatitis infection are the e antigens¹⁵, which recent studies suggest are a dissociated form of the core antigen¹⁶.

Here we report the cloning in bacteria and structural analysis of DNA from hepatitis B virus (HBV). We have defined the location of the viral surface antigen gene in the restriction endonuclease cleavage map of the cloned HBV genome. Nucleotide sequence of an 892-base pair region encoding the surface antigen gene allows us to deduce the complete amino acid sequence of this protein.

Cloning of DNA from hepatitis B virus Dane particles

DNA from Dane particles (prepared from pooled sera received from Merck, Sharp and Dohme) was labelled with ³²P-dATP and ³²P-dCTP using the endogenous DNA polymerase reaction¹⁰, and purified as described by Landers *et al.*¹¹. Several different DNA preparations were cleaved by the restriction endonuclease *Eco*RI into a single fragment of approximately 3,200 base pairs and by *Bam*HI into two fragments of approximately 2,100 and 1,100 base pairs (results not shown). For cloning, the Dane particle DNA was digested with *Bam*HI endonuclease and separately with *Eco*RI, ligated to the respective sites of pBR322 (ref. 17) and pBR325 (ref. 18) and used to transform *Escherichia coli* χ 1776. Ampicillin-resistant colonies derived from the DNA ligated into the *Bam*HI site were further screened for their sensitivity to tetracycline¹⁷ and the size of the plasmids they contained¹⁹. Similarly, recombinant colonies derived from the DNA ligated into the *Eco*RI site of pBR325 were screened for their sensitivity to chloramphenicol¹⁸ and by analysis of their plasmids¹⁹.

Putative recombinant plasmids were isolated and examined by agarose gel electrophoresis after digestion by *Eco*RI and *Bam*HI. One clone from the *Eco*RI experiment (named pHBV-3200) was found to contain a hybrid plasmid with an inserted DNA fragment of approximately 3,200 base pairs, the reported size of linearised viral DNA²⁰. A clone from the *Bam*HI

experiment (named pHBV-2100) was found to contain a plasmid with a 2,100-base pair insert. The identities of the inserts within pHBV-3200 and pHBV-2100 were verified by cleavage of the plasmids with *Eco*RI and *Bam*HI and hybridisation of the fragments with ³²P-labelled Dane particle DNA (Fig. 1). Physical mapping of both inserted fragments with restriction endonucleases showed that pHBV-2100 contains a *Bam*HI-derived insert also present in pHBV-3200.

Location and sequence analysis of the fragment containing the HBV surface antigen gene

A physical map containing the cleavage sites for several restriction endonucleases was prepared from pHBV-3200. This map has been obtained by a combination of enzyme digestions and DNA sequence analysis (Fig. 2). No restriction sites were found for *Xma*I, *Sst*I, *Kpn*I, *Hpa*I, *Xho*I and *Sac*I endonucleases. The map is similar, but not identical, to those recently described by others for cloned hepatitis viral DNA^{21–23}. For example, the map reported by Charnay *et al.*²³ shows an extra *Bam*HI site in the region containing the surface antigen gene. Examination of the nucleotide sequence in this region (Figs 3, 4) shows that the sequence GGATCA coding for the amino acids Gly-Ser may be converted to a *Bam*HI site GGATCC by a change of one base in the third position of the codon, which would therefore not result in a change in the amino acid sequence.

The HBV surface antigen gene was located by extensive nucleotide sequence analysis of the entire cloned viral genome to identify the sequence coding for the 19 amino acids present at the amino-terminus of the protein¹³. The location of the gene within the physical map of the cloned viral DNA is shown in Fig. 2.

The nucleotide sequence of an 892-base pair region encoding this gene was determined as summarised in Fig. 5 legend. The sequence, including the restriction enzyme cleavage sites, is presented in Fig. 3. The translation of this sequence in one of the phases (Fig. 4) predicts precisely the sequence of the 19 N-terminal amino acids of the proteins P1 and P2, except for residue 15, in which the DNA sequence predicts a leucine instead of the serine. Re-examination of the original data of Peterson and Vyas has shown that residue 15 is indeed leucine, and that the original report was incorrect. Furthermore, the amino acid sequencing data up to residue 31 is completely compatible with the amino acid sequence predicted by the DNA, with the exception of residue 24 (arginine), in which insufficient amounts of the phenylthiohydantoin amino acid were formed for analysis (D. Peterson, personal communication). The reported C-terminal sequence, Val-Tyr-Ile¹³, is in phase from 204 amino acid residues towards the 3' RNA terminus just before the ochre termination codon, UAA. The polypeptide encoded by these sequences is 226 amino acids long and has a MW of 25,398, in satisfactory agreement with the mass for P1 (22–24,000) determined by SDS gel electrophoresis¹². The amino acid composition also agrees very closely with that reported for this protein¹³. Note also the relatively high content of proline (10.2%), tryptophan (5.8%), aromatic amino acids (Trp, Tyr, Phe) (15.0%) and hydrophobic amino acids (Val, Ile, Leu) (26.6%).

Because of the prevalence of intervening sequences in eukaryotic genes (see ref. 24 and refs therein), it is not possible to presume the colinearity of a gene with the amino acid sequence of the protein product. However, there is no evidence for an intervening sequence in the surface antigen gene, as the molecule predicted by the DNA sequence closely approximates the characteristics of the isolated surface antigen P1. Any intervening sequence(s) would have to be small (<150 bases); most intervening sequences in structural genes are longer. The N-terminal and C-terminal ends of the molecule are in phase, thus any intervening sequence must also maintain the phase. Furthermore, preliminary studies suggest that the number of major peptides resolved from trypsin-treated aminoethylated

P1 agrees quite well with those predicted from the amino acid sequence derived from the gene (D. Peterson, personal communication). Further studies on P1 or its mRNA should resolve this issue decisively. In the meantime, it seems justified to assume colinearity of the gene with the mRNA for purposes of examination of the primary structure of P1. The high concentration of proline residues and their dispersion throughout the molecule preclude the possibility of a high content of α -helix. Indeed, according to the Fasman rules²⁵, there is only approximately 4% α -helix. However, there is a significant amount of β form (~30%) and also more than 10% β -turns. Thus, the molecule probably has a globular configuration. Perhaps the most striking primary structural feature of the molecule is a 19-amino acid long hydrophobic region. The sequences flanking this region are particularly rich in proline and cysteine residues. This hydrophobic region may be the site for intermolecular interaction of P1 monomers, and the abundant cysteine residues could make intra- and intermolecular linkages to form the surface coat of the virus. Because of the likelihood that P1 is glycosylated, we searched for possible sites of glycosylation at asparagine residues (Asn-X-Ser/Thr)²⁶. There are three such sites at amino acid positions 3, 59 and 146.

Directions for future study

Although P1 and P2 are the major peptides, five or six other proteins of larger size (up to ~100,000 MW) have been observed in surface antigen preparations¹². One of these, P6 (~72,000 MW), is sometimes present as a major component¹². This family of molecules cross-reacts antigenically and hence must be related structurally. Whether they represent aggregates of P1/P2 or are distinct molecules is not known. We have not detected nucleotide sequences encoding similar amino acid sequences in the other regions of the viral genome. The possibility that they are formed by processing from a single peptide precursor should also not be overlooked. However, the DNA sequence suggests that it is unlikely that a precursor could extend from the C-terminus because several other termination codons exist in the 3' RNA direction. Furthermore, the nucleotide sequence which could code for a protein of 190 amino acids begins with the methionine codon AUG, 41 bases after the termination codon of the surface antigen gene (Fig. 5). Screening of the sequence of the entire virus suggested that this may be the gene coding for the major core protein. Further structural analysis of the core protein must be carried out before a

*Bam*HI *Hin*FI *Mbo*II *Mn*II

GGATCCCAGAGTCAGGGGTCTGTATCTTCCTGCTGGTGGCTCCAGTTCAGGAACAGTAAACCCTGCTCCGAATATTGCCTCTCACATCTC
CCTAGGGTCTCAGTCCCCAGACATAGAAGGACGACCACCGAGGTCAAGTCTTGTCAATTTGGGACGAGGCTTATAACGGAGAGTGTAGAG

*Mbo*I

*Tac*I *U*I *Hin*FI *U*I

GTCAATCTCCGCGAGGACTGGGGACCTGTGACGAACATGGAGAACATCACATCAGGATTCTAGGACCCCTGCTCGTGTACAGGCGGG
CAGTTAGAGCGCTCCTGACCCCTGGGACACTGCTTGTACCTCTTGTAGTGTAGTCTTAAGGATCTTGGGACGAGCACAATGTCCGCC

*Mn*II *Ava*2 *Ava*2

*Hin*DI *Hin*FI *Hin*FI *Hin*FI *Mbo*I

GTTTTTCTTGTGACAAGAATCCTCACAAATACCGCAGAGTCTAGACTCGTGGTGGACTTCTCTCAATTTCTAGGGGGATCTCCCGTGTG
CAAAAAGAACAACTGTTCTTAGGACTGTTATGGCGTCTCAGATCTGAGCACCACCTGAAGAGAGTTAAAGATCCCCCTAGAGGGCACAC

*Mn*II *Xba*I

*Bal*I *Mn*II *Hph* *Mn*II *Mn*II *Apy*I

TCTTGGCCAAAATTCGCAGTCCCCAACCTCCAATCACTACCAACCTCCTGTCTCCAATTTGTCTGGTTATCGCTGGATGTGTCTGCG
AGAACCCTTTTAAAGCGTCAGGGGTGGAGGTTAGTGAGTGGTTGGAGGACAGGAGGTTAAACAGGACCAATAGCGACCTACACAGACGC

*Hae*III *Eco*RII

*Mn*II *Bbr* *Mn*II *Mbo*II *Mbo*I *Mn*II

GCGTTTTATCATATTCTCTTCATCTGCTGCTATGCCTCATCTTCTTATTGGTTCTTCTGGATTATCAAGGTATGTTGCCCGTTTGTCC
CGCAAATAGTATAAGGAGAAGTAGGACGACGATACGGAGTAGAAGAATAACCAAGAAGACCTAATAGTCCATACAACGGGCAACAGG

*Mbo*II

*Eco*RII *Mbo*II *U*I *Hin*FI *Mn*II

TCTAATTCAGGATCAACAACAACAGTACGGGACCATGCAAAACCTGCACGACTCCTGCTCAAGGCAACTCTATGTTTCCCTCATGTTG
AGATTAAGGTCTAGTTGTTGTTGGTCATGCCCTGGTACGTTTTGGACGTGCTGAGGACGAGTTCCTGTTGAGATACAAAGGAGTACAAC

*Apy*I *Ava*2

*Eco*RII *Asu*I *Mn*II

CTGTACAAAACCTACGGATGGAATTCACCTGTATTCCCATCCCATCGTCTGGGCTTTTCGCAAAATACCTATGGGAGTGGGCCTCAGT
GACATGTTTTGGATGCCTACCTTTAAGTGGACATAAGGGTAGGGTAGCAGGACCGAAAGCGTTTTATGGATACCCTCACCCGGAGTCA

*Apy*I *Hae*III

*Alu*I

CCGTTTTCTTGGCTCAGTTTACTAGTGCCATTTGTTTCACTGGTTCGTAGGGCTTTCCCCACTGTTTGGCTTTTCACTATATGGATGAT
GGCAAAGAGAACCAGTCAAATGATCAGGTAAACAAGTACCAAGCATCCCGAAAGGGGGTGACAAACCGAAAGTGGATATACCTACTA

*Asu*I *Hin*FI

GTGGTATTGGGGGCCAAGTCTGTACAGCATCGTGAGTCCCTTTATACCGCTGTTACCAATTTTCTTTTGTCTCTGGGTATACATTTAAAC
CACCATAACCCCGGTTTCAACATGTCGTAGCACTCAGGGAATATGGCGACAATGGTTAAAAGAAAACAGAGACCCATATGTAAATTTG

*Hae*III

*Mbo*I

CCTAACAAAACAAAAGATGGGGTTATTCCCTAACTTCATGGGCTACATAATTGGAAGTTGGGGAACTTTGCCACAGGATC
GGATTGTTTTGTTTTTCTACCCCAATAAGGGATTTGAAGTACCCGATGTATTAACCTTCAACCCCTTGAACGGTGTCTCTAG

Fig. 3 Nucleotide sequence of the HBV surface antigen gene and adjacent regions. Sequence analysis was carried out by the method of Maxam and Gilbert³⁴ as outlined in Fig. 4. The enzyme *Mbo*I is an isoschizomer of *Sau*3A. U1 indicates *Sau*₉₆I sites.

G GAU CCC AGA GUC AGG GGU CUG UAU CUU CCU GCU GGU GGC UCC AGU UCA GGA ACA GUA AAC CCU GCU CCG AAU AUU GCC UCU CAC AUC

1 10
met glu asn ile thr ser gly phe leu gly pro leu leu val leu gln ala
UCG UCA AUC UCC GCG AGG ACU GGG GAC CCU GUG ACG AAC AUG GAG AAC AUC ACA UCA GGA UUC CUA GGA CCC CUG CUC GUG UUA CAG GCG

20 30 40
gly phe phe leu leu thr arg ile leu thr ile pro gln ser leu asp ser trp trp thr ser leu asn phe leu gly gly ser pro val
GGG UUU UUC UUG UUG ACA AGA AUC CUC ACA AUA CCG AGU CUA GAC UCG UGG UGG ACU UCU CUC AAU UUU CUA GGG GGA UCU CCC GUG

50 60 70
cys leu gly gln asn ser gln ser pro thr ser asn his ser pro thr ser cys pro pro ile cys pro gly tyr arg trp met cys leu
UGU CUU GGC CAA AAU UCG CAG UCC CCA ACC UCC AAU CAC UCA CCA ACC UCC UGU CCU CCA AUU UGU CCU GGU UAU CGC UGG AUG UGU CUG

80 90 100
arg arg phe ile ile phe leu phe ile leu leu leu cys leu ile phe leu leu val leu leu asp tyr gln gly met leu pro val cys
CGG CGU UUU AUC AUA UUC CUC UUC AUC CUG CUG CUA UGC CUC AUC UUC UUA UUG GUU CUU CUG GAU UAU CAA GGU AUG UUG CCC GUU UGU

110 120 130
pro leu ile pro gly ser thr thr thr ser thr gly pro cys lys thr cys thr thr pro ala gln gly asn ser met phe pro ser cys
CCU CUA AUU CCA GGA UCA ACA ACA ACC AGU ACG GGA CCA UGC AAA ACC UGC ACG ACU CCU GCU CAA GGC AAC UCU AUG UUU CCC UCA UGU

140 150 160
cys cys thr lys pro thr asp gly asn cys thr cys ile pro ile pro ser ser trp ala phe ala lys tyr leu trp glu trp ala ser
UGC UGU ACA AAA CCU ACG GAU GGA AAU UGC ACC UGU AUU CCC AUC CCA UGC UCC UGG GCU UUC GCA AAA UAC CUA UGG GAG UGG GCC UCA

170 180 190
val arg phe ser trp leu ser leu leu val pro phe val gln trp phe val gly leu ser pro thr val trp leu ser ala ile trp met
GUC CGU UUC UCU UGG CUC AGU UUA CUA GUG CCA UUU GUU CAG UGG UUC GUA GGG CUU UCC CCC ACU GUU UGG CUU UCA GCU AUA UGG AUG

200 210 220 226
met trp tyr trp gly pro ser leu tyr ser ile val ser pro phe ile pro leu leu pro ile phe phe cys leu trp val tyr ile OC
AUG UGG UAU UGG GGG CCA AGU CUG UAC AGC AUC GUG AGU CCC UUU AUA CCG CUG UUA CCA AUU UUC UUU UGU CUC UGG GUA UAC AUU UAA

ACC CUA ACA AAA CAA AAA GAU GGG GUU AUU CCC UAA ACU UCA UGG GCU ACA UAA UUG GAA GUU GGG GAA CUU UGC CAC AGG AUC

Fig. 4 The amino acid sequence of the HBV surface antigen gene. The nucleotide sequence presented in Fig. 3 was translated in one reading frame so as to correspond to the known first 19 amino acids of the N-terminus of the HBV surface antigen¹³.

definitive assignment can be made.

It is possible that a putative precursor could extend from the N-terminus. In certain instances, promoter sites in eukaryotic genes may be identified from the DNA sequence. Hogness and Goldberg (personal communication) have postulated that the canonical sequence TATAAATA may interact with RNA polymerase, as *E. coli* polymerase interacts with the Pribnow box²⁷. Initiation of transcription usually occurs 23 ± 1 nucleotides from this site. We have been unable to find this sequence in the 200 nucleotides immediately preceding the N-terminal methionine. However, more distant variations on the Hogness-Goldberg sequence occur (D. S. Hogness, personal communication; E. Ziff, personal communication and ref. 28). For example, the sequence TATATT is found 184 bases in the 5' RNA direction from the initiator methionine codon of the surface antigen gene and also occurs in the IgG2 light chain²⁹. However, there are no specific data and the site of initiation of transcription and the structure of the initial gene product of the HBV surface antigen gene therefore remain in doubt.

Tiollais and coworkers have defined the single-stranded regions in the DNA of the virus²⁰. From their map, it is clear that

single-stranded regions of DNA extend into the surface antigen gene, but probably not as far as the initiation of transcription. These observations localise the coding strand of the gene on the full-length DNA strand.

The availability of abundant quantities of hepatitis B viral DNA through molecular cloning procedures will allow a comprehensive study of its gene products and the pathology of the virus. The present work enables specific studies on the expression of the gene coding for the major surface antigen to be carried out. The availability of this gene may also provide an alternative means for production of the antigen which in turn can be used as a vaccine against hepatitis B infection.

We thank Drs D. Peterson and G. Vyas for helpful discussions and for making their unpublished data available to us, Drs J. Birnbaum, R. Hirschman, M. Hilleman and A. Tytell for the Dane particle preparation and for helpful discussions, and Drs J. T. Yang and G. R. I. Chang for consultations on the putative structure of the surface antigen protein. This research was supported by a grant from Merck Sharpe and Dohme. H.M.G. is an Investigator of the Howard Hughes Medical Institute.

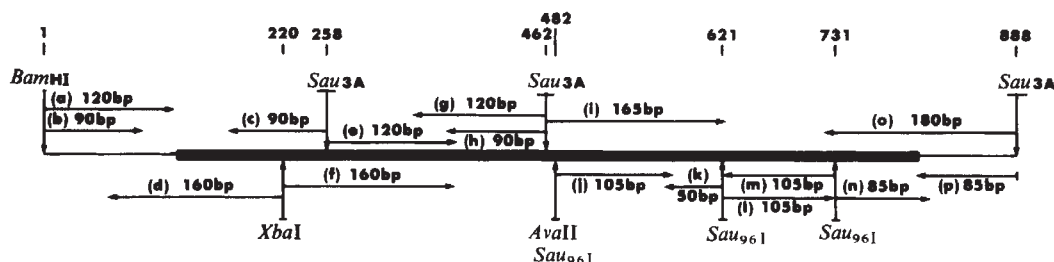


Fig. 5 Sequence strategy for the 889-base pair fragment containing the HBV surface antigen gene. Nucleotide residues are numbered in the direction from 5' to 3' in the message strand, beginning at the *Bam*HI site. The solid bar represents the 678-base pair region coding for the protein. Only the restriction sites used as starting points for sequencing are shown. For labelling, DNA was treated with 5 μ g of bacterial alkaline phosphatase (Worthington) at 37 °C for 60 min in 20 mM Tris-HCl (pH 8.0), phenol extracted and treated with 10 units of T4 polynucleotide kinase (Boehringer) at 37 °C for 30 min in a reaction mixture containing 50 mM glycine (pH 9.5), 10 mM MgCl₂, 10 mM dithiothreitol, 0.1 mM spermidine and 0.001 mM [γ -³²P]ATP ($\sim 4,000$ Ci mmol⁻¹). Fragments labelled only at one end were isolated by gel electrophoresis after either digestion with a restriction enzyme or strand separation. DNA sequence was carried out by the method of Maxam and Gilbert³⁴. Arrows indicate the direction and number of base pairs sequenced in each experiment. Sequences were determined in both strands or in duplicate for all but 50 residues. In this 50-base pair region the data were very reliable. (Sequence data are available from the authors on request.)

Received 28 June; accepted 30 July 1979.

1. Krugman, S. in *Viral Hepatitis. A Contemporary Assessment of Etiology, Epidemiology, Pathogenesis and Prevention* (eds Vyas, G. N., Cohen, S. N. & Schmid, R.) 3–10 (Franklin Institute Press, Philadelphia, 1978).
2. Feinstone, S. M., Kapikian, A. Z. & Purcell, R. H. *Science* **182**, 1026–1028 (1973).
3. Provost, P. J. *et al. Proc. Soc. exp. Biol. Med.* **148**, 532–539 (1975).
4. Deinhardt, F., Holmes, A. W., Capps, R. B. & Popper, H. J. *exp. Med.* **125**, 673–688 (1967).
5. Mascoli, C. C. *et al. Proc. Soc. exp. Biol. Med.* **142**, 276–282 (1973).
6. Provost, P. J. & Hilleman, M. R. *Proc. Soc. exp. Biol. Med.* **160**, 213–221 (1979).
7. Dane, D. S., Cameron, C. H. & Briggs, M. *Lancet* **i**, 695–698 (1970).
8. Blumberg, B. S. *Science* **197**, 17–25 (1977).
9. Destmyer, J., Ray, M. G., Bradburne, A. F. & Alexander, J. J. in *Viral Hepatitis. A Contemporary Assessment of Etiology, Epidemiology, Pathogenesis and Prevention* (eds Vyas, G. N., Cohen, S. N. & Schmid, R.) 459–463 (Franklin Institute Press, Philadelphia, 1978).
10. Hruska, J. F., Clayton, D. A., Rubenstein, J. L. & Robinson, W. S. *J. Virol.* **21**, 666–672 (1977).
11. Landers, T. A., Greenberg, H. B. & Robinson, W. S. *J. Virol.* **23**, 368–376 (1977).
12. Gerin, J. L. & Shi, J. W. K. in *Viral Hepatitis. A Contemporary Assessment of Etiology, Epidemiology, Pathogenesis and Prevention* (eds Vyas, G. N., Cohen, S. N. & Schmid, R.) 147–153 (Franklin Institute Press, Philadelphia, 1978).
13. Peterson, D. L., Chien, D. Y., Vyas, G. N., Nitecki, D. & Bond, H. E. in *Viral Hepatitis. A Contemporary Assessment of Etiology, Epidemiology, Pathogenesis and Prevention* (eds Vyas, G. N., Cohen, S. N. & Schmid, R.) 569–573 (Franklin Institute Press, Philadelphia, 1978).
14. Shi, J. W. K. & Gerin, J. L. *J. Virol.* **21**, 347 (1977).
15. Magnus, L. O. in *Viral Hepatitis. A Contemporary Assessment of Etiology, Epidemiology, Pathogenesis and Prevention* (eds Vyas, G. N., Cohen, S. N. & Schmid, R.) 177–182 (Franklin Institute Press, Philadelphia, 1978).
16. Takahashi, K. *et al. J. Immun.* **122**, 275–279 (1979).
17. Bolivar, F. *et al. Gene* **2**, 95–113 (1977).
18. Bolivar, F. *Gene* **4**, 121–136 (1978).
19. Barnes, W. M. *Science* **195**, 393–394 (1977).
20. Summers, J. A., O'Connell, A. & Millman, I. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4597–4601 (1975).
21. Burrell, C. J., Mackay, P., Greenaway, P. J., Hofschneider, P. H. & Murray, K. *Nature* **279**, 43–47 (1979).
22. Sninsky, J. J., Siddiqui, A., Robinson, W. J. & Cohen, S. N. *Nature* **279**, 346–348 (1979).
23. Charnay, P., Pourcel, L. A., Fritsch, A. & Tiollais, P. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2222–2226 (1979).
24. Robertson, M. S. *et al. Nature* **278**, 370–372 (1979).
25. Chou, P. Y. & Fasman, G. D. *Biochemistry* **13**, 21–222, 222–245 (1974).
26. Winkler, R. J. in *Hormonal Proteins and Peptides* (ed. Li, C. H.) 1–15 (Academic, New York, 1973).
27. Pribnow, D. *Proc. natn. Acad. Sci. U.S.A.* **72**, 784–788 (1975).
28. Ziff, E. B. & Evans, R. M. *Cell* **15**, 1463–1475 (1978).
29. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1485–1489 (1978).
30. Ullrich, A. *et al. Science* **196**, 1313–1319 (1977).
31. Clewell, D. B. & Helinski, D. R. *Proc. natn. Acad. Sci. U.S.A.* **62**, 1159–1166 (1969).
32. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
33. Rigby, P. W. J., Dieckman, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237–251 (1977).
34. Maxam, A. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).

letters

Discovery of IR bursts from Liller I/MXB1730–333

THE first detection of IR bursts from the object¹ known as Liller I which has been identified² with the rapid X-ray burster MXB1730–333 is reported here. The rapid burster is the only one of its kind that has been extensively studied in the X-ray region^{3–5} and some of its known characteristics can be summarised as follows. When it is active it gives several thousand X-ray bursts per day. It operates in two modes: mode I (usually in March and September every year), mode II (usually in April and October every year). In mode I it gives large X-ray bursts with energy in the range 10^{39} – 10^{40} erg and also small bursts with energy in the range of 10^{38} – 10^{39} erg; while in mode II it primarily emits bursts with energy typically close to 10^{39} erg. These bursts are called type II on the basis of the constant character of their spectrum during the decay phase. The rapid burster occasionally also gives bursts known as type I which are characterised by the softening of the X-ray spectrum during the decay phase. The energy in these bursts is usually in excess of 10^{39} erg and their frequency is about one every few hours.

Liller¹ searched for an optical counterpart of the rapid burster and found a highly reddened compact cluster with a red magnitude $m_r = 21$ in the error box of the rapid burster⁶. This has been identified as a globular cluster by Kleinmann *et al.*⁷, with the help of IR observations. The identification of this cluster with the rapid burster was supported by Doxsey *et al.*² who obtained a more precise position for the X-ray burster. Kleinmann *et al.*⁷ observed the globular cluster continuously for about an hour on 29 May 1976 and did not detect any variation greater than 10% of the total IR flux. However, the rapid burster may have ceased activity in April 1976.

Two of us (K.M.V.A. and S.M.C.)⁸ have proposed a model for the X-ray emission from the rapid burster based on accretion of matter onto the poles of a rotating magnetic neutron star. More recently (unpublished data), the cyclotron emission of electrons in the accreting column above the poles of the magnetised

neutron star has been considered and emission in the IR and optical regions suggested; these calculations have prompted us to undertake the present experiment. The optical emission from the rapid burster may, however, be difficult to detect because of the large interstellar extinction^{1,7} of $A_v = 11 \pm 1$ mag. On the other hand, the interstellar extinction in the IR is very small which makes it possible to observe the radiation.

The observations were made on the 1-m telescope of the Indian Institute of Astrophysics at Kavalur (lat. 12° N), India on the night of 4–5 April 1979 between 21.29 and 00.10 UT. At the Cassegrain focus a liquid-nitrogen cooled photometer was used with a set of IR filters, three apertures and a Fabry mirror imaging the primary of the telescope on a $0.5 \text{ mm} \times 0.5 \text{ mm}$ InSb detector. The incoming beam was $f/20$ giving the scale of 10 arc s mm^{-1} at the focal plane. For the present measurement we used a 2 mm aperture corresponding to 20 arc s field of view. The incoming beam was chopped alternately on the source and a neighbouring part of the sky with a throw of 20 arc s by a tertiary mirror inclined at 45° to the incident beam with a frequency of 16 Hz. The servo-controlled square wave gave a duty cycle of 75%. The Liller I signal was detected with the phase sensitive detector followed by the low pass filter with a time constant of 0.6 s. In the case of standard stars, the phase sensitive detector

Table 1 IR bursts from Liller I/MXB1730–333

Burst No.	Time of occurrence (UT)	Rise time (s)	Duration (s)	FWHM (s)	Peak luminosity (10^{37} erg s^{-1} in $0.3 \mu\text{m}$ interval at $1.6 \mu\text{m}$)	Burst energy (10^{38} erg in $0.3 \mu\text{m}$ interval at $1.6 \mu\text{m}$)
1	21 h 33 min 00 s	2	37	12	2.2	3.0
2	22 41 24	2	36	12	1.9	2.0
3	23 45 06	3	36	12	2.0	2.6
4	23 47 30	3	24	9	1.7	1.5
5	00 03 00	2	36	10	2.2	2.5
6a	00 04 54	2	20	6	2.3	1.7
6b	00 05 12	2	19	6	1.6	0.9

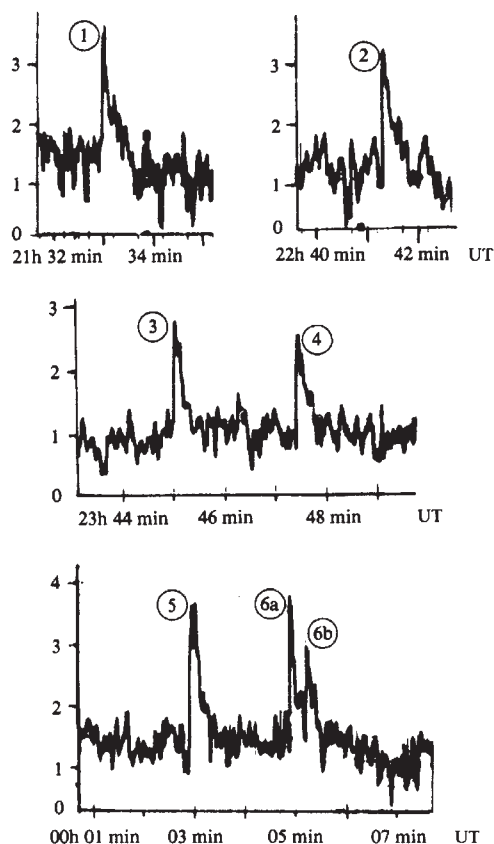


Fig. 1 Six IR bursts from the source Liller I/MXB1730-333, the rapid burster. Note the composite structure of burst number 6; the two components are designated a and b. The ordinate numbers are in units of $10^{-16} \text{ W cm}^{-2} \mu\text{m}^{-1}$.

output was linearly integrated for 6 s. The measurements reported here were made with an H-filter (λ peak $1.6 \mu\text{m}$, half width $0.3 \mu\text{m}$, maximum transmission $\sim 75\%$). Several standard IR sources were observed before the observation of Liller I.

The steady IR flux from Liller I was observed and its magnitude calculated as $m_H = 7.5 \pm 0.4$ using a nearby star⁹ IRC-30317 for calibration. Kleinmann *et al.*⁷ have reported $m_H = 6.4$ for Liller I using 24 arc s beam size and 90 arc s chopper throw. The present measurement has (1) a smaller beam and (2) a smaller chopper throw. Taking the flux distribution of Liller I as given by Kleinmann *et al.*⁷ we estimate that our measurement will be higher by $m_H = 0.54$. After accounting for (1) and (2) above and subtracting 0.54 from our m_H value, it is still higher by 0.56 ± 0.4 than the flux value given by Kleinmann *et al.*⁷.

Six IR bursts detected during the 2.5 h observing time are shown in Fig. 1. The details of the time of occurrence, rise time, duration, FWHM, peak intensity and energy in the bursts are summarised in Table 1. Kleinmann *et al.*⁷ have given interstellar extinction in the K-band $A_K = 1.1$ mag for Liller I. Assuming an inverse wavelength dependence for extinction, we estimate $A_H = 1.4$ mag. While calculating peak luminosity and energy we have subtracted $A_H = 1.4$ mag from observed m_H values. The energy in the bursts is estimated by integrating the area under the burst profiles and assuming a distance of 10 kpc to the object³. Note the close similarity of the various characteristics of the successive IR bursts. They seem to rise rapidly in a time period of nearly 2 s and decay gradually over time of about 30 s. Note that burst number 6 is composed of two bursts, the characteristics of which are separately given in Table 1. The occurrence of burst numbers 3, 4 and 5, 6 in rapid succession (~ 150 s) may be significant.

The remarkably close similarity of the rise time, duration and gradual decay of intensity of the IR bursts with the corresponding characteristics of type-I X-ray bursts^{3,4} strongly indicate the

association between the two. The low rate of occurrence of the IR bursts accords with this association.

The observed IR emission from the bursts is unlikely to be of black body type. If a region of 2 light seconds in size is assumed then the peak burst luminosity implied a brightness temperature of $7 \times 10^7 \text{ K}$ at $1.6 \mu\text{m}$ if the radiation were thermal. This is greater than the X-ray burst temperature and would imply a much greater total luminosity.

The observations of the rapid burster at the X-ray and IR wavelengths suggest that there could be emission of burst radiation at all intermediate wavelengths. It is likely that in spite of large interstellar extinction the optical bursts from the rapid burster would be detected. It will also be interesting to observe other X-ray burst sources at different IR wavelengths.

We thank Dr M. K. V. Bappu for providing time on the Kavalur telescope at short notice and for valuable discussions, also K. S. B. Manian for technical assistance, and the staff of the Kavalur Observatory. We thank Professors R. R. Daniel, D. Lal, S. P. Pandya, B. V. Sreekantan and M. S. Vardya for encouragement.

P. V. KULKARNI
N. M. ASHOK

Physical Research Laboratory,
Ahmedabad 380009, India

KRISHNA M. V. APPARAO
S. M. CHITRE

Tata Institute of Fundamental Research,
Bombay 400005, India

Received 18 April; accepted 27 June 1979.

1. Liller, W. *Astrophys. J. Lett.* **213**, L21-L23 (1977).
2. Doxsey, R. *et al. Astrophys. J. Lett.* **221**, L53-L55 (1978).
3. Lewin, W. H. G. in *Advances in Space Exploration*, Vol. 3, (in the press).
4. Hoffman, J. A., Marshall, H. L. & Lewin, W. H. G. *Nature* **271**, 630-633 (1978).
5. Marshall, H. L., Ulmer, M. P., Hoffman, J. A., Doty, J. & Lewin, W. H. G. *Astrophys. J.* **227**, 555-562 (1979).
6. Lewin, W. H. G. *IAU Circ. No.* 2922 (1976).
7. Kleinmann, D. E., Kleinmann, S. G. & Wright, E. L. *Astrophys. J. Lett.* **210**, L83-L86 (1976).
8. Apparao, K. M. V. & Chitre, S. M. *Astrophys. Space Sci.* (in the press).
9. Neugebauer, G. & Leighton, R. B. *Two-Micron Sky Survey* (NASA Sp-3047, 1969).

Possible chemical impact of planetary lightning in the atmospheres of Venus and Mars

THE importance of lightning as a source of trace gases in the Earth's atmosphere has long been recognised. In 1827, von Liebig proposed lightning as a source of atmospheric fixed nitrogen¹ and recent investigations have supported this hypothesis²⁻⁷. In the Earth's pre-biological atmosphere lightning may have produced organic molecules, such as HCN, which led to the evolution of life⁸. After life evolved, but before photosynthesis and biological nitrogen fixation developed, lightning may have provided an abiotic source of organic C and fixed N to the growing biota^{9,10}. The presence of cloud layers on Venus¹¹ and synoptic scale dust storms on Mars¹² suggest that lightning may also occur on these planets, and indeed there have been reports of lightning being detected on Venus by American and Soviet spacecraft. The implications for atmospheric chemistry of lightning on Venus and Mars are discussed here.

The production of gases by an electrical discharge involves high-temperature chemical reactions in and around the discharge channel¹³; the air around the channel is heated by the radially propagated shock wave. As the temperature, T , of the heated air decreases at some rate characterised by a cooling time, $\tau_T(T)$, a species X may attain a final concentration in

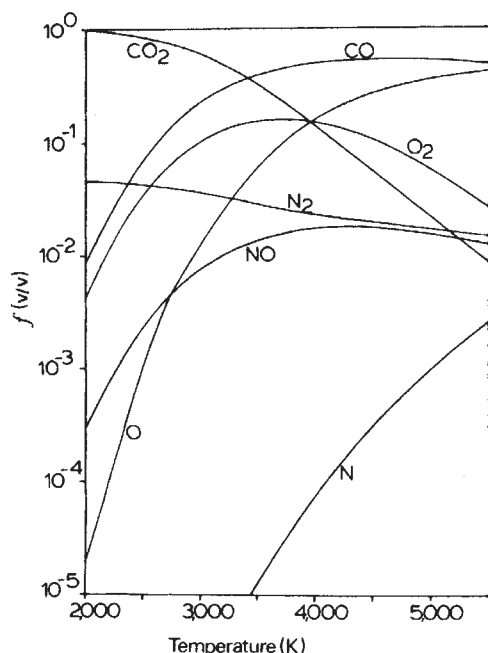


Fig. 1 Equilibrium mixing ratios of CO_2 , N_2 , O_2 , O , NO and N as a function of temperature in heated Cytherean air. To obtain these results, the equilibrium abundances of the 24 species previously listed⁷ plus S , S_2 , S_2O , SO , SO_2 , SO_3 , COS , CS_2 and H_2SO_4 were determined as a function of temperature using the thermochemical data from the JANAF tables¹⁴. Ambient conditions were chosen for an altitude of 55 km, the lower cloud region; ambient $T = 310$ K (ref. 15), ambient $\text{CO}_2 = 0.66$ atm, $\text{N}_2 = 0.032$ atm, $\text{SO}_2 = 1.4 \times 10^{-4}$ atm, $\text{O}_2 = 4 \times 10^{-5}$ atm, $\text{H}_2\text{O} = 4 \times 10^{-4}$ atm, and $\rho = 1.2$ kg m^{-3} (ref. 16).

excess of its ambient value, thus leading to a net production of X . The final concentration of X is approximated by the equilibrium concentration $f_x^e(T_F(X))$ (v/v) at $T_F(X)$, the 'freeze-out' temperature for X . T_F is given by the temperature at which

$$\tau_x(T_F(X)) = \tau_T(T_F(X)) \quad (1)$$

where τ_x is the chemical lifetime of species X . When $T > T_F$, then $\tau_x < \tau_T$ and the chemical reactions are sufficiently rapid to keep X in thermochemical equilibrium. However, for $T < T_F$, $\tau_x > \tau_T$ and chemical reactions are too slow to adjust to the rapidly decreasing T ; X , therefore, 'freezes-out' with mixing ratio $f_x^e(T_F(X))$. While a lower abundance in X is favoured thermodynamically at low T , the kinetics are too slow for readjustment. P , the net yield of X due to this process, is approximated by⁷

$$P(X) \approx \frac{f_x^e(T_F)M(E_0, T_F(X))}{E_0} \quad (\text{molecules J}^{-1}) \quad (2)$$

where M is the number of molecules per metre heated to or above T_F in the region where X is being produced, and E_0 is the discharge energy (in J m^{-1}). Generally, T_F is between 1,000 and 5,000 K.

Figure 1 illustrates the thermochemical equilibrium concentrations of several species of interest for air in the cloud layer of Venus heated to temperatures of 2,000–5,500 K, for $A = \rho/\rho_0 = 1$ where ρ is the mass density within the heated air and ρ_0 is the ambient mass density. The variation of martian equilibrium composition with temperature is quite similar to that of Venus. For both planets we find significant enhancements of CO , O_2 , NO and O for $T > 2,000$ K, indicating that these species may be produced by lightning on Venus and Mars. Production of these gases by lightning could have significant implications for the

chemistry and evolution of the atmospheres of Venus and Mars and needs, therefore, to be investigated.

Whether lightning does, in fact, produce trace species depends on the cooling characteristics of the discharge; an estimate of the source rates requires knowledge of the energy properties and the frequency of planetary lightning. Thus, a reliable prediction of the chemical impact of lightning on Venus and Mars will have to await appropriate observations. However, we estimate the source rates by assuming that lightning of Venus is similar to that on Earth, as an illustration.

The source rates of CO , O_2 , NO , and O on Venus with Earth-like lightning are estimated using an earlier reported method⁷; $\tau_T(T)$ and $M(T, E_0)$ are calculated from Lin's¹⁷ differential equations describing a cylindrical shock wave. Yields predicted by this method for NO , N_2O and CO in discharges in present atmospheric conditions have compared favourably with laboratory experiments^{4,7,18}. The shock wave differential equations are solved with a specific heat ratio, $\gamma = 1.16$, appropriate to a CO_2 atmosphere at temperatures between 1,500 K and 5,000 K¹⁹. As these equations indicate that, for the temperature region of interest, $1 \leq A = \rho/\rho_0 \leq 10$, f_x^e and τ_x are determined for $A = 1$ and 10 as limiting values. Finally, as CO_2 , which comprises about 95% of the ambient atmosphere, is significantly dissociated at temperatures above 2,000 K, stoichiometrically it follows that virtually all O atoms in our system must originate from CO_2 dissociation and we therefore require that $P(\text{CO}) = P(\text{NO}) + 2P(\text{O}_2) + P(\text{O})$.

Using the earlier method

$$\tau_T(T) \sim \left(\frac{1}{T(R)} \frac{dT(R)}{dt} \right)^{-1} = \frac{5.7 \times 10^{-5} E_0^{1/2}}{\rho_0^{1/2} T} \quad (\text{s}) \quad (3)$$

for $\gamma = 1.16$, where R is the shock front radius, and t is time. For $E_0 \sim 10^5$ J m^{-1} , as on Earth²⁰, and $\rho_0 = 1.2$ kg m^{-3} at 55 km, the lower portion on Venus' cloud layer, $\tau_T \sim (5-7) \times 10^{-6}$ s for $2,500 \leq T \leq 5,000$ K, as depicted in Fig. 2. Figure 2 also shows τ_{CO} , τ_{O_2} and τ_{NO} . Using equation (1) to determine T_F , Fig. 2 indicates that CO freezes-out first; for $A = 1$, $T_F(\text{CO}) = 5,500$ K with $f_{\text{CO}}^e(5,500 \text{ K}) = 0.5$ and for $A = 10$, $T_F(\text{CO}) = 3,500$ K with

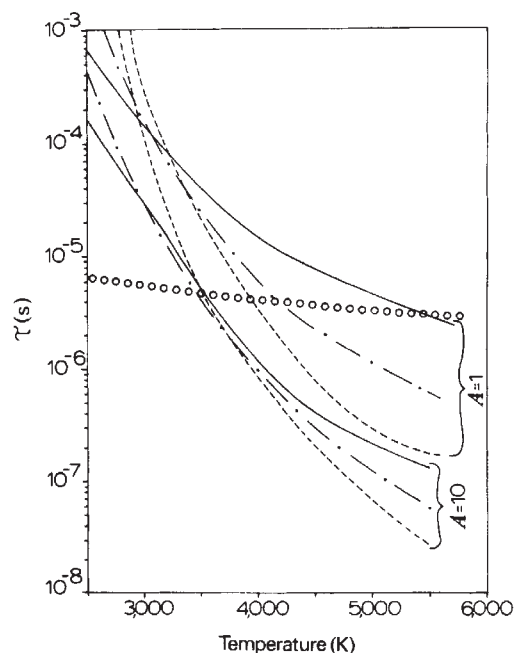


Fig. 2 Temperature dependence of the cooling time, τ_T ($\circ$), and chemical lifetimes; τ_{CO} (—), τ_{O_2} (---); and τ_{NO} (···) for $A = 1$ and 10 where $A = \rho/\rho_0$. The chemical lifetimes were determined from densities taken from the thermochemical equilibrium calculations and the appropriate kinetic data^{21,22}.

Table 1 Estimated production rates for Earth-like lightning on Venus

Species	$f_z(T_F(X))^*$ (v/v)		$P(X)$ (molecules J^{-1})		$S(X)^{\dagger}$ (molecules $cm^{-2} s^{-1}$)	
	$A = 1$	$A = 10$	$A = 1$	$A = 10$	$A = 1$	$A = 10$
CO	0.50	0.26	1.4×10^{17}	1.1×10^{17}	1.4×10^{10}	1.1×10^{10}
O ₂	0.16	0.12	4.5×10^{16}	5×10^{16}	4.5×10^9	5×10^9
NO	0.018	0.013	5×10^{15}	6×10^{15}	5×10^8	6×10^8
O	0.16	0.018	4.5×10^{16}	0.8×10^{16}	4.5×10^9	0.8×10^9

* For cooling time characterised by $E_0 = 10^5 J m^{-2}$, $\rho = 1.2 kg m^{-3}$.

† For $10^{-7} W cm^{-2}$ dissipated by lightning.

$f_{CO}(3,500 K) = 0.26$. For $\gamma = 1.16$, Lin's¹⁷ equations yield

$$M(T_F, E_0) = \frac{1.55 \times 10^{21}}{T_F} E_0 \quad (\text{molecules } m^{-1}) \quad (4)$$

and, substituting equation (4) into equation (2), $P(CO) \sim (1.1-1.4) \times 10^{17}$ molecules J^{-1} .

The yields of O₂, NO and O are controlled by the CO kinetics. As CO₂ dissociation is the only source of O atoms, O₂, NO and O cannot be produced outside the CO-producing region. In the CO-producing region, the equilibrium abundances and chemical lifetimes of O₂, NO and O below $T_F(CO)$ may be determined from the thermochemical calculations with the CO abundance set at $f_{CO}(T_F(CO))$. The final number of O₂, NO and O molecules or atoms produced is then given by their respective freeze-out abundances multiplied by $M(T_F(CO), E_0)$, the numbers of molecules heated to or above $T_F(CO)$. We find that O₂ and NO freeze-out at approximately the same temperature; for $A = 1$, $T_F(O_2) \sim T_F(NO) \sim 4,000 K$, and for $A = 10$, $T_F(O_2) \sim T_F(NO) \sim 3,300 K$. The final concentration of O, which freezes-out at a lower T , is determined by conservation of oxygen. Table 1 summarises the estimated final abundances of the species and their yields, and also indicates the resulting global source rates, S , obtained by assuming that lightning on Venus, as on Earth, dissipates about $10^{-7} W cm^{-2}$.

Before this study it was believed that O, O₂ and CO were produced exclusively above the clouds by photochemistry^{23,24}. Downward transport is thought to supply the lower atmosphere with O₂ and CO and thereby fuel the conversion of S and SO₂ to H₂SO₄-H₂O particles and SO₂ to liquid S, components of the clouds of Venus²⁵⁻²⁹. Photochemical models²³ estimate downward fluxes at 62 km of 1.4×10^{12} CO molecules $cm^{-2} s^{-1}$ and 5.5×10^{11} O₂ molecules $cm^{-2} s^{-1}$. Comparison with the numbers in Table 1 indicates that, if lightning on Venus were similar to that on Earth, it would contribute a small, additional, local source of CO and O₂ to augment the upper atmospheric source. However, if lightning on Venus were 100 times more energetic or frequent, and this is certainly a possibility, the CO and O₂ discharge sources would be comparable with the downward transport source and thus could have a key role in maintaining the clouds. If in the 55 km region SO₂ is about 100 p.p.m. (ref. 16), the major sink for lightning-produced O is probably $O + SO_2 + M \rightarrow SO_3 + M$, further augmenting SO₄²⁻ production.

While we predict 2-3 times more CO production than O₂ production by lightning, Oyami *et al.*¹⁶ found at least 60 times more O₂ than CO in the lower atmosphere of Venus. However, note that the abundance of an atmospheric constituent is a function of both the loss and source terms. In the terrestrial atmosphere, for example, CO is considerably more reactive than O₂ and thus, while the O₂ source is only 100 greater than that of CO^{30,31}, the O₂ abundance is 10⁶ greater than the CO concentration. While a similar effect may cause the apparent enhancement of O₂ relative to CO on Venus, complete photochemical model calculations which include all the relevant sources and sinks are needed to explain the observed composition of the atmosphere of Venus.

The production of NO by lightning may result in the formation of HNO₂ and HNO₃ and thus imply the trace presence of NO₂⁻ and NO₃⁻ in the clouds. The NO₂⁻ and NO₃⁻ are probably removed in precipitation and may either be recycled back to N₂ or remain on the surface. For the NO yield obtained, the time to

consume atmospheric N₂ is about 5 billion years and thus, in the absence of recycling, roughly equal amounts of nitrogen would reside in the atmosphere and lithosphere. A similar effect on Mars could help to explain the apparent deficiency of nitrogen on this planet relative to Earth. However, on Mars we find rather low yields for lightning with energies of $10^5 J m^{-2}$ and cooling rates of about $10^{-6} s$; the thin martian atmosphere leads to slow chemical reaction rates, high freeze-out temperatures, and low chemical yields. For lightning on Mars to lead to significant chemical production cooling times of 0.01-1.0 s are required which, according to equation (3), would imply extremely energetic discharges.

We find that lightning, if present, may produce CO, O₂, NO and O in the atmospheres of Venus and Mars. Calculations for lightning on Venus indicate that this process could conceivably influence the atmospheric nitrogen budget and the sulphur chemistry that maintains the Venus cloud layer. A final determination of the chemical role of planetary lightning, however, will have to await systematic observations of the presence and properties of electrical discharges on Venus and Mars.

The National Astronomy and Ionosphere Center is operated by Cornell University under contract with the NSF. This work was supported partly by the NSF under Grant ATM77-20179 to W.L.C.

Note added in proof: Since completion of this work, documentation of the evidence for lightning on Venus gathered by American and Soviet spacecraft has appeared^{32,33}, further emphasising the potential importance of planetary lightning.

W. L. CHAMEIDES

Department of Physics and Astronomy,
University of Florida,
Gainesville, Florida 32611

J. C. G. WALKER

National Astronomy and Ionosphere Center,
Arecibo, Puerto Rico 00612

A. F. NAGY

Department of Atmospheric
and Oceanic Science,
University of Michigan,
Ann Arbor, Michigan 48109

Received 5 April; accepted 5 July 1979.

1. von Liebig, J. *Am. Chem. Phys.* **35**, 329-333 (1827).
2. Noxon, J. F. *Geophys. Res. Lett.* **3**, 463-465 (1976).
3. Zipf, E. C. & Dubin, M. *Trans. Am. geophys. Un.* **57**, 965 (1976).
4. Chameides, W. L., Stedman, D. H., Dickerson, R. R., Rusch, D. W. & Cicerone, R. J. *J. Atmos. Sci.* **34**, 143-149 (1977).
5. Tuck, A. F. *Q. J. R. Met. Soc.* **102**, 749-755 (1976).
6. Griffing, G. W. *J. geophys. Res.* **82**, 943-950 (1977).
7. Chameides, W. L. *Nature* **277**, 123-125 (1979).
8. Miller, S. L. & Urey, H. C. *Science* **130**, 245-251 (1959).
9. Yung, Y. L. & McElroy, M. B. *Science* **203**, 1002-1004 (1979).
10. Chameides, W. L. & Walker, J. C. G. in preparation.
11. Knollenberg, R. G., Hansen, J., Ragent, B., Martonchik, J. & Tomasko, M. *Space Sci. Rev.* **20**, 329-354 (1977).
12. Gierasch, P. J. *Rev. Geophys. Space Phys.* **13**, 790-871 (1975).
13. Zel'dovich, Y. B. & Raizer, Y. P. *Physics of Shock Waves and High-Temperature Hydrodynamics Phenomena* (Academic, New York, 1966).
14. Stull, D. R. & Prophet, H. *JANAF Thermochemical Tables* (NSRDS-NBS 37, US Govt. Printing Office, Washington, D.C., 1971).
15. Seiff, A. *et al. Science* **203**, 787-790 (1979).
16. Oyami, V. I., Carle, G. C., Woeller, F. & Pollack, J. B. *Science* **203**, 802-805 (1979).
17. Lin, S. C. *J. appl. Phys.* **25**, 54-57 (1954).
18. Levine, J. S., Hughes, R. E., Chameides, W. L. & Howell, W. E. *Geophys. Res. Lett.* **6**, 557-559 (1979).
19. Zucrow, M. I. & Hoffmann, J. D. *Gas Dynamics*, Vol. 1 (Wiley, New York, 1976).
20. Uman, M. A. *Lightning* (McGraw-Hill, New York, 1969).
21. Baulch, D. L., Drysdale, D. D., Horne, D. G. & Lloyd, A. C. *Evaluated Kinetic Data for High Temperature Reactions*, Vol. 2 (Chemical Rubber Co., Cleveland, 1973).
22. Hampson, R. F. & Garvin, D. (eds) *Chemical Kinetic and Photochemical Data for Modelling Atmospheric Chemistry* (NBS Tech. Note 866, NBS, Washington, D.C., 1975).
23. Sze, N. D. & McElroy, M. B. *Planet. Space Sci.* **23**, 763-786 (1975).
24. Liu, S. C. & Donahue, T. M. *Icarus* **24**, 148-156 (1975).
25. Hapke, B. & Nelson, R. J. *atmos. Sci.* **32**, 1212-1218 (1975).
26. Young, A. T. *Icarus* **32**, 1-26 (1977).
27. Young, A. T. *Geophys. Res. Lett.* **6**, 49-50 (1979).
28. Prinn, R. G. *Science* **182**, 1132 (1973).
29. Wofsy, S. C. & Sze, N. D. *Atmospheres of Earth and the Planets*, Vol. 51 (ed. McCormac, B. M.) (Reidel, Dordrecht, 1975).
30. Walker, J. C. G. *Evolution of the Atmosphere* (Macmillan, New York, 1977).
31. Seiler, W. *Tellus* **26**, 117-135 (1974).
32. Taylor, W. L., Scarf, F. L., Russell, C. T. & Brace, L. H. *Nature* **279**, 614-616, 1979.
33. Ksanfomaliti, L. V. *et al. Electrical Discharges in the Venus Atmosphere* (D-272, Academy of Sciences, Space Research Institute, Moscow, USSR, 1979).

Photoinduced electron transfer across a water-oil boundary as a model for redox reaction separation

THE creation of artificial solar energy conversion and storage systems that mimic the photosynthetic pathway has evoked great interest in recent years¹. One approach to the photolysis of water involves the mediation of two photosystems in the generation of reduced and oxidised species that are the active components in the decomposition of water^{2,3}. However, homogeneous (aqueous) solutions of these components suffer from the basic limitation that the reducing and oxidising agents can react with each other and thus no net reaction can be observed. Several kinds of synthetic 'photosynthetic membranes' have been suggested as a means of separating the two redox units and overcoming these fundamental difficulties³. Recently, photosensitised electron transfer across vesicle walls has been demonstrated and suggested as a means for generating oxidising and reducing agents in separate water compartments⁴. We report here a photochemical electron transfer across the interface of a water-in-toluene microemulsion, and propose this system as a model for the separation of oxidised and reduced

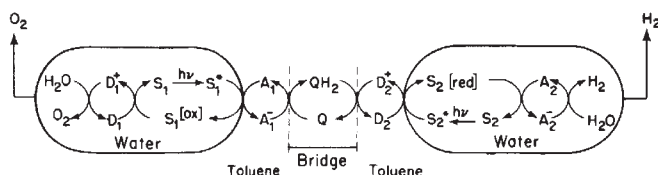


Fig. 1 General scheme for water photodecomposition using two half cells of water-in-toluene microemulsions.

species. It is well known that surfactant molecules aggregate to reversed micelles in organic solvents⁵. Reversed micelles entrap water to form 'water pools' in a continuous oil phase. The proposed general model for the compartmentalisation of two water phases and its utilisation in the photolysis of water is shown in Fig. 1.

The model system involves the generation of an oxidised donor (D_1^+) and a reduced acceptor (A_2^-) in the water droplets of two separate half cells using the two sensitizers S_1 and S_2 respectively. The complementary redox agents of this process (A_1^- and D_2^+) are confined to the organic phase, so that back reactions are inhibited. The two half cells are bridged by electron and proton carriers to regenerate A_1 and D_2 . Regeneration of D_1 and A_2 is coupled to oxidation and reduction of water, so these must be charge storage catalysts^{2,3}. Thus, all the components, except water, are recycled.

In principle, the electron transfer process of one half cell can be divided into two distinct parts: (1) from the water to the interphase, (2) from the interphase into the continuous organic solution. Electron transfer from the water phase to the interphase was investigated in a microemulsion in which the 'water pools' included EDTA as the donor, and tris(2,2'-bipyridine)ruthenium(II) ($\text{Ru}(\text{bipy})_3^{2+}$) as the photosensitizer. The acceptor, 1,1'-dihexadecyl-4,4'-bipyridinium chloride (HV^{2+}), is expected to be located at the interphase boundary as a result of its amphiphilic structure. The microemulsion was prepared by adding 0.15 ml of 0.3 M ammonium EDTA solution ($\text{pH} = 8.5$) and 21 μl of 10^{-2} M $[\text{Ru}(\text{bipy})_3]\text{Cl}_2$ solution to 2.9 ml of toluene containing 220 mg of dodecylammonium propionate as surfactant (0.3 M with respect to the surfactant). The mixture was vortex stirred to clarity. Into the resulting solution 2 mg of HV^{2+} were dissolved by vortex stirring. The microemulsion solution was transferred into a cuvette of 1-cm pathlength, and an inert atmosphere in the cell was generated by evacuation followed by nitrogen bubbling through the solution for 15 min. The contents of the sealed cuvette were stirred while being

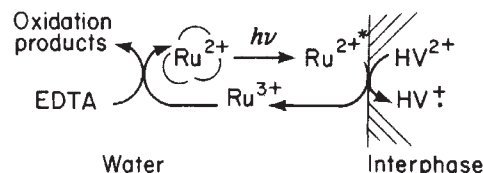


Fig. 2 Electron transfer from the aqueous phase to an acceptor in the interphase.

illuminated with blue light (incident photon flux $\sim 10\text{--}16 \times 10^{-7}$ einstein s^{-1}). The formation of $\text{HV}^{\bullet+}$ was followed spectroscopically⁶. After 4 min of illumination 75% of the parent HV^{2+} had been reduced ($\phi_{\text{max}} = 1.3 \pm 0.4\%$). Introducing air into the cuvette reoxidised $\text{HV}^{\bullet+}$ to HV^{2+} and revealed that no change in the original $\text{Ru}(\text{bipy})_3^{2+}$ concentration occurred. Furthermore, the fact that the mole ratio $\text{HV}^{2+}:\text{Ru}(\text{bipy})_3^{2+}$ was 13:1 emphasises that the photosensitizer acted as a catalyst and was recycled during the $\text{HV}^{\bullet+}$ formation.

The size of the water droplets^{5,7} ensures numerous collisions of the sensitizer with the interface during its excited state lifetime ($\sim 0.6 \mu\text{s}$)⁸. Hence, the results are rationalised by the electron transfer from photoexcited $\text{Ru}(\text{bipy})_3^{2+}$ to the interphase located acceptor HV^{2+} . The resulting $\text{Ru}(\text{bipy})_3^{3+}$ is reduced by EDTA and thus the photosensitizer is recycled. Therefore, electron transfer from a donor in the water phase to an acceptor in the interphase has been demonstrated (see Fig. 2).

However, in order to simulate the general scheme proposed in Fig. 1, electron transfer into the continuous organic phase is required. For this, an acceptor that is originally located in the interphase, while its reduced form is extracted from this region, into the bulk organic phase, seems to be desirable. Furthermore, coupling of this acceptor with a second acceptor present in the continuous organic phase should regenerate the acceptor at the interface. Benzylnicotinamide (BNA^+) seems to fulfill these requirements, as it is expected to concentrate at the interface, while its reduced form should migrate into the bulk organic solution⁹. Hence, we have investigated the photosensitised reduction of the dye, 4-dimethylaminoazobenzene (dye) present in the continuous organic phase, mediated by BNA^+ at the interface and $\text{Ru}(\text{bipy})_3^{2+}$ in the water droplets. As in the previous experiment, EDTA was used as donor. The microemulsion system was prepared as follows. 230 mg of dodecylammonium propionate and 3.7 mg of BNA^+ (4.8×10^{-3} M) were dissolved in 2.9 ml of a 5×10^{-5} M solution of 4-dimethylaminoazobenzene in toluene. 0.1 ml of ammonium EDTA solution (0.3 M, $\text{pH} = 8.5$) and 3 μl of $\text{Ru}(\text{bipy})_3^{2+}$ solution (10^{-2} M) were added to the toluene solution. The resulting microemulsion solution was illuminated under an inert atmosphere with blue light. The progress of the electron transfer was followed by the disappearance of the dye absorption at $\lambda = 402 \text{ nm}$ ($\epsilon = 22,000 \text{ mol}^{-1} \text{ cm}^{-1}$). The change in the dye concentration as a function of illumination time is shown in Fig. 3. After 4 min of illumination, 80% of the dye had been reduced

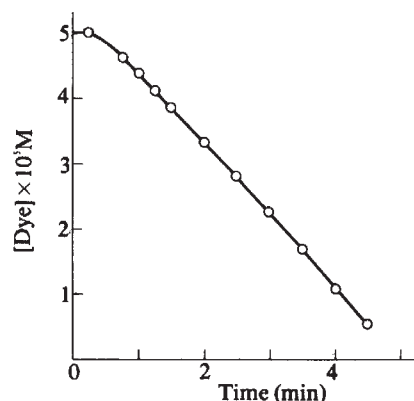


Fig. 3 The reduction of 4-dimethylaminoazobenzene as a function of illumination time.

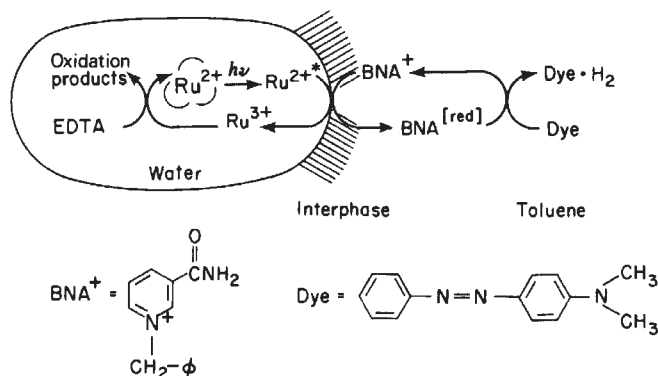


Fig. 4 Cyclic mechanism for electron transfer across the interface.

($\phi_{\text{max}} = 0.13 \pm 0.04\%$). The reduced dye product (Dye H_2) was identified as the corresponding hydrazo derivative. On the introduction of I_2 or H_2O_2 , the reduction product Dye H_2 could be reoxidised to the diazo-dye¹⁰. At the end of the dye photoreduction, the $\text{Ru}(\text{bipy})_3^{2+}$ concentration seemed to be unchanged. The experimental mole ratio of the dye: $\text{Ru}(\text{bipy})_3^{2+}$ (5:1) indicates the photocatalytic activity of the ruthenium complex. Furthermore, in control experiments excluding any of $\text{Ru}(\text{bipy})_3^{2+}$, EDTA, or BNA^+ , no reduction of the dye could be observed on illumination. Hence, all the components of the system are crucial in the reduction process. The fact that BNA^+ is required in the process emphasises that it acts as a mediating agent in photoreduction of the dye.

These results are rationalised by a cyclic redox mechanism as shown in Fig. 4. The electron transfer from photoexcited $\text{Ru}(\text{bipy})_3^{2+}$ to BNA^+ is followed by the reduction of the dye dissolved in the continuous organic phase. The photosensitiser is in turn regenerated by the oxidation of EDTA. The thermodynamic balance of the net reaction, reduction of the dye by EDTA to the corresponding hydrazo derivative, is uphill in free energy ($\Delta G^\circ \sim 37$ kcal per mol of EDTA consumed)¹¹. Thus, the process represents a net storage of energy.

Our results demonstrate that an endoergic redox reaction across a water-oil interface has been carried out. The compartmentalisation of the redox reaction may serve as a key solution in the photolysis of water. Further steps, involving reversible acceptor-donor couples separated by this interface, as well as the introduction of adequate agents for the oxidation and reduction of water, are being investigated.

This work was supported by the Division of Chemical Sciences Office of Basic Energy Sciences of the US Department of Energy under contract W-7405-Eng-48. I.W. thanks the Weizmann Institute of Science for a Chaim Weizmann Postdoctoral Fellowship.

ITAMAR WILLNER*
WILLIAM E. FORD
JOHN W. OTVOS
MELVIN CALVIN

Laboratory of Chemical Biodynamics,
Lawrence Berkeley Laboratory,
University of California,
Berkeley, California 94720

Received 6 April; accepted 27 June 1979.

* To whom reprint should be addressed.

- Balzani, V., Maggi, L., Manfrin, M. F., Bolletta, F. & Gleria, M. *Science*, **189**, 852-856 (1975).
- Bolton, J. R. *Science* **202**, 705-711 (1978).
- Calvin, M. *Acc. chem. Res.* **11**, 369-374 (1978).
- Ford, W. E., Otvos, J. W. & Calvin, M. *Nature* **274**, 507-508 (1978).
- Fendler, J. H. & Fendler, F. J. *Catalysis in Micelles and Macromolecular Systems* (Academic, New York, 1975).
- Steckman, E. & Kuwana, T. *Ber. Bunsenges* **78**, 253-259 (1974).
- Lim, Y. Y. & Fendler, J. H. *J. Am. chem. Soc.* **100**, 7490-7494 (1978).
- Van Houten, J. & Watts, R. J. *J. Am. chem. Soc.* **98**, 4853-4858 (1976).
- Nakamura, K., Ohno, A., Yasui, S. & Oka, S. *Tetrahedron Lett.* 4815-4818 (1978).
- Newbold, B. T. in *The Chemistry of the Hydrazo, Azo and Azoxy Groups* (ed. Patai, S.) Chap. 15 (Wiley, New York, 1975).
- Krasna, A. I. *Photochem. Photobiol.* **29**, 267-276 (1979).

Strong electric field effects on proton transfer between membrane-bound amines and water

DURING water demineralisation by electrodialysis much of the current is carried by hydrogen and hydroxyl ions^{1,2}. This is attributed to an effect known as water splitting, which occurs near those surfaces of the ion exchange membranes which are depleted of ions during current flow (see Fig. 1). I have found³ that water splitting only occurs in sulphonic acid membranes if certain impurities are present. But in contrast, it is an intrinsic property of anion exchange membranes in which the functional group is a tertiary alkyl amine. It is also intrinsic to membranes with quaternary ammonium ions if the surface groups have been converted to the tertiary form. An investigation of water splitting for the latter type of membrane (Negev Institute A)⁴ is described here. It will be shown that the effect could be due to a 1,000-fold increase in the forward rate constant for reaction (2), caused by the strong external electric field (10^7 - 10^8 V m⁻¹) at the interface.

The measurements were made in a cell with eight chambers (Fig. 2). The central two were separated by the anion exchange membrane under study. The others isolated the central chambers from the pH changes occurring at the electrodes. Water splitting originated in the surface layer of the anion exchange membrane, on the cathode side (see Fig. 1). The current carried by the hydrogen and hydroxyl ions (J_w), was calculated from the rate of p_H change of the central solutions.

Water splitting was only observed when the current exceeded a threshold. It then generally accounted for about one-third of the total current. The maximum value of J_w at room temperature exceeded 50 A m⁻². At higher temperatures J_w increased, saturating at a value of about 400 A m⁻² at 100 °C, in 10 mM and 0.1 M NaCl solutions. The measurements were reproducible to within 10%; however, the temperature at the surface of the membrane was uncertain over a wide range of currents because of local heating.

To appreciate the magnitude of the effect consider the production of H^+ and OH^- ions in a membrane from ordinary water dissociation. For a membrane 30 μm thick with an average water concentration (C_w) of 10 M the rate is 6×10^{-6} mol m⁻² s⁻¹. These ions recombine; however, even if they did reach the membrane boundaries the contribution to J_w would be only 0.6 A m⁻². Considering that water splitting only originates near the surface, the experimental value of 50 A m⁻² at room temperature is several orders of magnitude greater than what could possibly be attributed to ordinary water dissociation.

It is now shown that the reaction layer is in the space charge region at the surface of the membrane (see Fig. 1). We define boundaries $x = 0$, δ of the reaction layer according to $J_H = 0.99 J_w$ at $x = 0$ and $J_{OH} = 0.99 J_w$ at $x = \delta$. Outside this region reaction flow is negligible and $\mu_w + \mu_H + \mu_{OH} = 0$, where μ denotes electrochemical potential. The outward flow (ϕ) of water ions is balanced by an inward flow of water molecules; $C_w \gg C_H, C_{OH}$. Thus

$$\left| \frac{d\mu_w}{dx} \right| \ll \left| \frac{d\mu_H}{dx} \right| \quad \text{at } x = 0 \quad \text{and} \quad \left| \frac{d\mu_w}{dx} \right| \ll \left| \frac{d\mu_{OH}}{dx} \right| \quad \text{at } x = \delta$$

using $\phi = -C_u(d\mu/dx)$, where u denotes mobility. It follows that $d\mu_H/dx \approx -d\mu_{OH}/dx$ at $x = 0, \delta$. Therefore $C_H u_H / C_{OH} u_{OH} \approx 99$ at $x = 0$ and $C_{OH} u_{OH} / C_H u_H \approx 99$ at $x = \delta$. We assume that u_H / u_{OH} and $C_H \cdot C_{OH}$ have the same values in the aqueous pathways in the membrane as in free solution, that is 1.8 and 10^{-14} mol² l⁻² respectively. The average water concentration is 55 M in the external solution and about 10 M in the membrane interior, where it forms aqueous pathways. The calculated concentrations just outside the reaction layer, are therefore $C_H = 7.4 \times 10^{-7}$ M, near $x = 0$, and 1.4×10^{-9} M, near

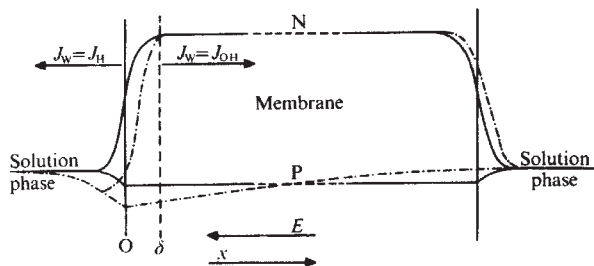
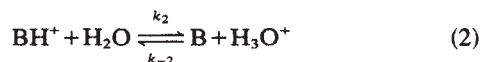
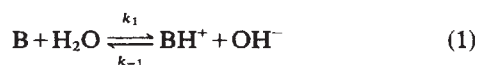


Fig. 1 Concentration profiles for positive (P) and negative (N) mobile ions, in an anion exchange membrane, at equilibrium (—) and during current flow (---).

$x = \delta$. These, and the values for C_{OH} , are negligible compared with the average concentration, that is ≈ 0.7 M, of ionisable groups in the membrane. Thus since such a large fraction of the total current is carried by the ions of water, it is inferred that the total concentration of mobile ions in the reaction layer is negligible compared with that of the fixed charge and that the reaction layer is therefore in the region of uncompensated fixed charge (Fig. 1).

A plausible mechanism which takes account of the tertiary amino groups (B) in the surface region is that the H_3O^+ and OH^- ions originate in protonation and deprotonation reactions that is



The times required for the H_3O^+ and OH^- ions to migrate out of the reaction layer under their gradients in electrochemical potential would need to be comparable or less than their recombination times with the amino groups. This condition seems to be fulfilled (unpublished work).

We shall estimate a lower limit for the pseudo-first-order rate constant k'_2 . Since the diffusion flow equals the integrated reaction flow

$$J_w < Fk_2[BH^+][H_2O]\delta \quad (3)$$

where $[BH^+]$ has been treated as being constant with x , and F is the Faraday constant. The change in electrical field intensity (E) across the reaction layer satisfies

$$E(0) - E(\delta) = \frac{F[BH^+]\delta}{\epsilon_0\epsilon_r} \quad (4)$$

using Poisson's equation, where ϵ_0 is the permittivity of free space and ϵ_r is the average dielectric constant. Using equations (3) and (4)

$$k'_2 > J_w / \epsilon_0\epsilon_r[E(0) - E(\delta)] \quad (5)$$

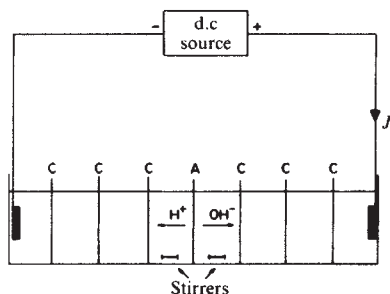


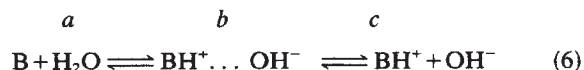
Fig. 2 The experimental system. The central chambers contained NaCl solutions of various concentrations. The other chambers were separated by cation exchange membranes and contained 0.1 M NaCl.

The upper limits for $[E(0) - E(\delta)]$ and ϵ_r are about 10^8 V m $^{-1}$ and 20 respectively. Thus to account for $J_w = 50$ A m $^{-2}$, k'_2 must exceed $2,800$ s $^{-1}$.

This estimate may be compared with that for the same amino group, but in the membrane interior. The functional group in the reaction layer seems to be the same as that in the Negev Institute tertiary alkyl amino ion exchange membranes³ which have an apparent pK_a of about 9.9 (unpublished work). Using $k_{-2} < 2.1 \times 10^{10}$ l mol $^{-1}$ s $^{-1}$ (ref. 5) we obtain $k'_2 > 2.5$ s $^{-1}$, for groups inside the membranes. This is less than a 1,000th of the value in the reaction layer.

Note that the major assumption underlying this result is that the great enhancement in the rate of water dissociation may be accounted for by reactions (1) and (2).

In considering how k'_2 might increase, the reactions are presented as occurring in two steps



where $BH^+ \dots OH^-$ and $B \dots H_3O^+$ are encounter pairs in the same solvent cage, k_{bc} and k_{cb} are the rate constants for the separation and encounter processes respectively and k_{ab} and k_{ba} are those for the chemical transformations. Applying the steady-state approximation to the encounter pair, the expressions for the overall rate constants are⁵

$$\bar{k} = \frac{k_{ab}k_{bc}}{k_{ba} + k_{bc}}; \quad \bar{k} = \frac{k_{cb}k_{ba}}{k_{ba} + k_{bc}} \quad (8)$$

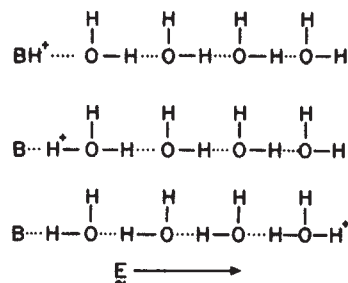


Fig. 3 Deprotonation of amino group and proton conduction along a chain of orientated water molecules.

Most of the difference in $\bar{k}$ between the membrane interior and the reaction layer would probably be due to a difference in $k_{bc}/(k_{ba} + k_{bc})$ rather than to k_{ab} , since the height of the potential barrier between states a and b is unlikely to alter by more than kT in the applied field. Therefore since the maximum value of $k_{bc}/(k_{ba} + k_{bc})$ is less than unity, the model requires that the ratio should be less than about 10^{-3} in the membrane interior.

From data for ammonia, the methylamines and triethylamine in water it seems likely that the reverse reactions are diffusion controlled⁶. For diffusion-controlled protolytic reactions in water k_{bc} is between 10^{10} and 10^{11} s $^{-1}$ and $k_{ba} > 10^{12}$ s $^{-1}$ (ref. 5). The constant k_{bc} is probably at least an order of magnitude smaller in the membranes because of the smaller ion diffusion coefficients⁷. However, the constant k_{ba} could again exceed 10^{12} s $^{-1}$, since the molecules may be correctly orientated for the reaction when the solvent cage is formed⁸. Thus the ratio $k_{bc}/(k_{ba} + k_{bc})$ could be less than 10^{-3} in the membrane interior, as required. We shall consider two ways by which it might be increased by the strong electric field in the reaction layer.

The first way resembles the second Wien effect. Onsager⁹ has shown for reaction (1), that k_{bc} , but not k_{cb} , depends on the strength of an external field. In reaction (2) the site in B, for its recombination with H_3O^+ , is a lone pair so that the charge product is effectively negative when the proton is sufficiently close to the reaction site. The dissociation of the protonated

amino group is, in this respect, analogous to that of a neutral group. It follows, from a consideration of Onsager's treatment, that k_{bc} would increase with the electric field intensity for reaction (2) also, although the magnitude of the increase is uncertain and depends on the variation of the interaction energy with the distance of separation and the role of the solvent in proton transfer¹⁰.

The second possibility involves the degree of alignment of the water molecules in the reaction layer. In free solution actual proton transfer is very fast and probably consists of quantum mechanical tunnelling. The rate-determining step seems to be the necessary reorientation of water molecules between successive transfers (structural diffusion)^{5,11}. It is conceivable that there is a bridge of water molecules in the reaction layer, already favourably orientated for proton transfer by the strong electric field, when reaction (2) occurs (Fig. 3). If this were the case then the rate-determining step in proton transfer would no longer be structural diffusion but quantum mechanical tunnelling through adjacent hydrogen bonds. The constant k_{bc} would then be much greater than in the membrane interior. It might even exceed k_{ba} , particularly if there were concerted proton transfer. Information relevant to this proposition should be obtainable from dielectric constant measurements for d.c. currents both below and above the threshold for water splitting.

Finally, note that since the electric field strength in the reaction layer is of the same order of magnitude as in biological membranes it is possible that proton transport in biological membranes involves water splitting and that the degree of protonation of amino groups there, depends on field strength. Such effects could be significant in several aspects of membrane function.

Part of this work was performed while I was on study leave at the Botany Department, Hebrew University, Jerusalem. I thank Professor F. de Körösy of the Negev Institute for Arid Zone Research, Beersheva for stimulating discussions.

R. SIMONS

School of Physics,
The University of New South Wales,
PO Box 1, Kensington,
New South Wales 2033, Australia

Received 4 April; accepted 22 June 1979.

1. Sonin, A. A. & Grossman, G. *J. phys. Chem.* **76**, 3996 (1972).
2. Shaffer, L. H. & Mintz, M. S. *Principles of Desalination*, 199 (ed. Spiegler, K. S.) (Academic, New York, 1966).
3. Simons, R. *Desalination* **28**, 41 (1979).
4. de Körösy, F. & Shorr, J. *Israel Pat. No.* 14720 (1962).
5. Eigen, M. *Angew. chem. Int. Edn.* **3**, 1 (1964).
6. Ralph, III, E. K. & Grunwald, E. *J. Am. chem. Soc.* **89**, 2963 (1967).
7. Simons, R. & Sajkewycz, N. *J. Membrane Biol.* **34**, 263 (1977).
8. Emerson, M. T., Grunwald, E. & Kromhout, R. A. *J. chem. Phys.* **33**, 547 (1960).
9. Onsager, L. *J. chem. Phys.* **2**, 599 (1934).
10. Albery, W. J. *Progress in Reaction Kinetics* **4**, 353 (1967).
11. Zundel, G. *The Hydrogen Bond*, 683 (eds. Schuster, P., Zundel, G., & Sandorfy, C.) (North-Holland, Amsterdam, 1976).

Natural atmospheric ^{14}C variation and the Suess effect

THE dilution of the atmospheric $^{14}\text{CO}_2$ concentration by large amounts of fossil-fuel derived CO_2 which do not contain any ^{14}C is commonly called the Suess effect. Its magnitude can be calculated with the same geochemical models as the global carbon cycle that also predict the future rise of atmospheric CO_2 to be caused by the combustion of fossil fuels. Validation of a CO_2 predictive model with the Suess effect ^{14}C data is important because these two phenomena have a common cause, and therefore register model responses at roughly the same frequencies. Measurements of the Suess effect yield values between -15% and -25% in $\Delta^{14}\text{C}$ (in 1950)¹, while different model predictions also cover about this range^{2,3}. The requirement that a model correctly reproduces the Suess effect becomes a strong constraint when the accuracy of the measurement is

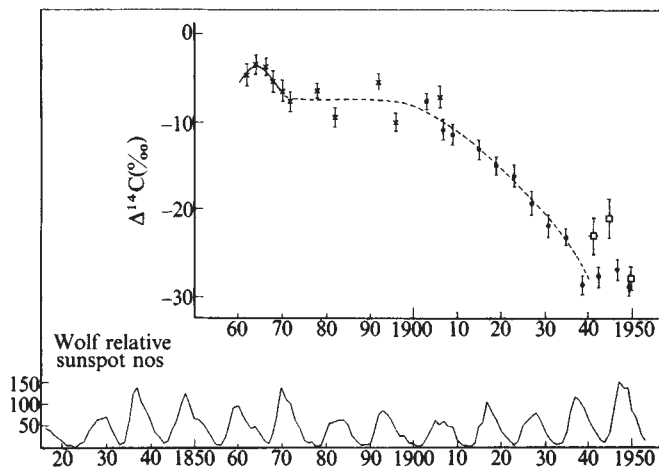


Fig. 1 $\Delta^{14}\text{C}$ in tree rings from Drenthe, The Netherlands: ●, *Quercus rubra* L. (Sleen, 52° 45' N, 6° 45' E); ×, *Quercus robur* L. (Sleen); □, *Quercus robur* L. (Norg, 53° 4' N, 6° 27' E). The line has been fitted by eye. The sunspot record has also been included for comparison.

improved to better than 2%. ^{14}C measurements in tree rings to an accuracy of 1.2% are reported here. The results indicate that the natural fluctuations of atmospheric ^{14}C as yet preclude determination of the Suess effect to the accuracy required by the models.

The design and the operating characteristics of the new set of proportional counters that were used for these measurements have been described elsewhere⁴. The samples were taken from three large mature oak trees from the province of Drenthe, The Netherlands, that had been felled by a severe windstorm in November 1972. The first 20 years of growth of each tree have been discarded. Only latewood was used from one of the trees, *Quercus rubra* L. which had very wide yearly rings. ^{14}C measurements were made on cellulose since this is the component of wood unlikely to contain carbon derived from other years of growth⁵. The amount of wood needed for every sample, before the cellulose extraction, was very large, ~200 g.

The results are listed in Table 1 and plotted in Fig. 1 as $\Delta^{14}\text{C}$ values defined by

$$\Delta^{14}\text{C} = \frac{A - A_{\text{ox}}}{A_{\text{ox}}} \times 1,000 \quad (\text{‰})$$

where A_{ox} is equal to $0.95 \times$ activity of NBS oxalic acid, corrected for age and fractionation to 1950 and -19.0% respec-

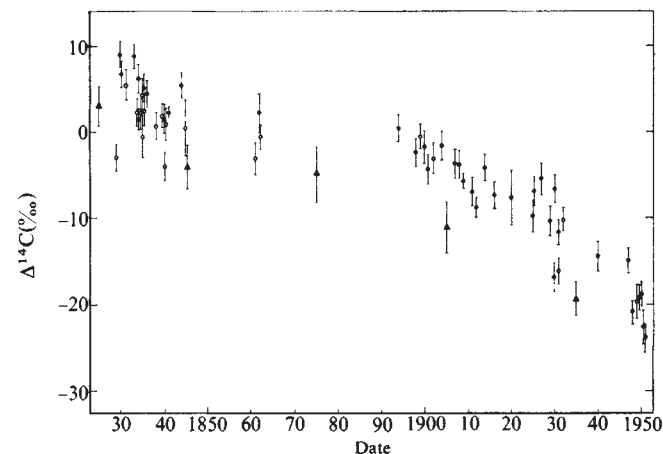


Fig. 2 Suess effect measurements by Lerman *et al.*¹: ●, Northern Hemisphere; ○, Southern Hemisphere. A few older results of de Vries, corrected for fractionation by Lerman *et al.*¹, are also included (▲).

Table 1 Radiocarbon content in tree rings calculated with a half life of 5,730 yr

Year of growth	Laboratory no.	$\delta^{13}\text{C}(\text{‰})$	$\Delta^{14}\text{C}(\text{‰})$
<i>a Quercus rubra</i> L. (only latewood)			
Post bomb period			
1972	GrN-8016	-23.51	474.6 ± 2.0
1971	8017	-23.88	498.5 ± 2.0
1967	8066	-23.68	637.3 ± 1.6
1961	8067	-24.86	229.4 ± 1.6
1959	8070	-22.55	277.9 ± 1.5
Pre bomb period			
1950	8069	-24.17	-29.0 ± 1.1
1947	8136	-23.74	-27.0 ± 1.3
1943	8137	-23.68	-27.9 ± 1.3
1939	8138	-24.09	-28.9 ± 1.1
1935	8139	-23.29	-23.3 ± 0.9
1931	8161	-24.05	-21.8 ± 1.3
1927	8162	-24.94	-19.3 ± 1.4
1923	8185	-25.62	-16.2 ± 1.3
1919	8192	-25.46	-15.0 ± 1.0
1915	8193	-25.53	-13.2 ± 1.2
1909	8198	-25.26	-11.5 ± 1.3
1907	8195	-25.24	-10.9 ± 0.9
1903	8196	-25.12	-7.7 ± 1.1
<i>b Quercus robur</i> L. (Sleen)			
1906	8364	-24.65	-7.1 ± 1.3
1895-1896	9003	-24.33	-9.7 ± 1.1
1892	8363	-24.26	-5.3 ± 0.9
1882	8362	-24.08	-9.5 ± 1.2
1878	8770	-24.18	-6.3 ± 1.0
1872	8199	-24.25	-7.7 ± 1.1
1870	8200	-24.37	-6.5 ± 1.2
1868	8201	-23.37	-5.3 ± 1.2
1866	8202	-24.77	-3.7 ± 0.9
1864	8203	-23.88	-3.4 ± 1.3
1862	8204	-24.11	-4.7 ± 1.2
<i>c Quercus robur</i> L. (Norg)			
1950	8295	-24.13	-28.1 ± 1.3
1945	8296	-24.39	-21.1 ± 2.2
1941-1942	8297	-24.37	-23.1 ± 2.1

tively, and A is the activity of the sample, also age and fractionation corrected from 1977 to the year of growth ($T_{1/2} = 5,730$ yr) and to -25.0‰ respectively.

One 11-yr period, 1862-72, could be closely followed because of an extended period of wide annual rings. The results agree with the dependence of atmospheric ^{14}C on the 11-yr solar cycle⁶, also depicted in Fig. 1. A precise magnitude for this dependence cannot be obtained until more solar cycles have been investigated and a longer term trend has been established for the 11-yr period to be fully explained.

Another striking feature is that the atmospheric ^{14}C level seems to remain nearly constant between 1939 and 1950. This result is in accordance with earlier measurements for this period by Damon, Long and Wallick⁶, who find an almost constant level around $\Delta^{14}\text{C} = -32\text{‰}$. During this 11-yr period a total of 16.6 G ton of carbon from fossil fuels has been released into the atmosphere⁷, which should have resulted in a 4-6% decrease of the atmospheric $^{14}\text{C}/\text{C}$ ratio during this decade. The observed constancy even occurs during a period of high solar activity when the natural ^{14}C production by cosmic rays is expected to be below average. One is tempted to speculate that the increased ^{14}C production which apparently cancels the Suess effect in these years is due to solar flares. An exceptionally large flare could conceivably raise the atmospheric ^{14}C level by several % (ref. 8). Increases in the cosmic ray intensity associated with flares was first reported in 1946 and 1949^{9,10}.

Furthermore it is not clear from the measurements which ^{14}C concentration should be chosen as the preindustrial level. There

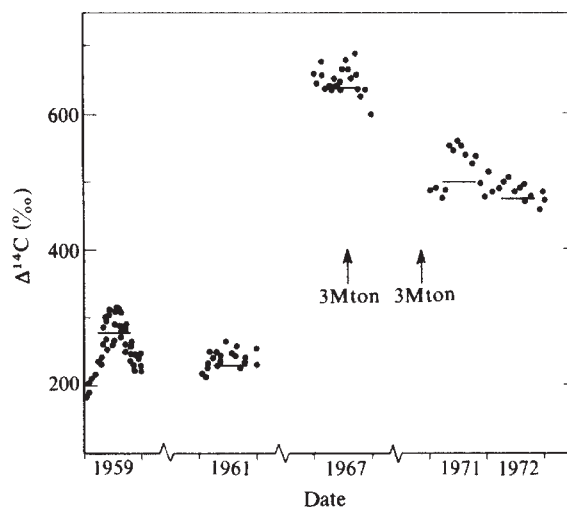


Fig. 3 Comparison of direct atmospheric ^{14}C measurements and tree rings. The 1959 and 1961 data are taken from Münnich¹¹ and Östlund *et al.*¹². The other data are from Nydal *et al.*¹³. The dates and estimated magnitudes in megatons TNT of two Chinese thermonuclear explosions are shown by arrows. The tree ring $\Delta^{14}\text{C}$ values are represented by horizontal bars through the summer half year.

does not seem to be such a level. Therefore, the Suess effect should be calculated as the difference between the measured time series and an extrapolation of the natural ^{14}C level from 1850 to 1950. This extrapolation would have to be based on an established relationship between atmospheric ^{14}C and the solar record.

When our results are compared with previous Suess effect measurements (Fig. 2) by Lerman, Mook and Vogel¹, our results are seen to be lower than the latter by a constant amount of 5-7%.

As our trees were located in a rural environment, at a typical distance of 150-200 km from and almost surrounded by the industrialised regions of North Western Europe, a regional fossil fuel contamination exceeding the average global effect cannot be fully excluded. This contamination cannot be estimated using the comparison with the results of Lerman *et al.* because of the large scatter of the older data. Comparing a number of recent growth rings with direct atmospheric ^{14}C measurements (Fig. 3) also does not lead to a definite quantitative conclusion. These data are also very scattered and some of them may be influenced by leakage from the stratosphere of radioactive debris from two recent Chinese thermonuclear explosions. Nevertheless, the fact that the tree $\Delta^{14}\text{C}$ values seem to be consistently on the low side of the possible range would suggest that the 1950 ring should be at most 4% lower in $\Delta^{14}\text{C}$ than the 1950 atmosphere.

A regional contamination is expected to increase almost in proportion with the global Suess effect. A comparison between the results of Lerman *et al.* and our data do not reveal conclusive evidence with respect to a contamination effect in our case.

Our observations of the short term variability of the natural ^{14}C level, however, remain equally valid regardless of the existence of a regional contamination effect.

We thank Dr C. D. Keeling for a helpful discussion of the results. This work has been supported by ZWO (Netherlands Organization for the advancement of Pure Research).

P. P. TANS*
A. F. M. DE JONG
W. G. MOOK

Isotope Physics Laboratory,
University of Groningen,
The Netherlands

Received 20 April; accepted 5 July 1979.

*Present address: Scripps Institute of Oceanography, University of California at San Diego, La Jolla, California 92093.

1. Lerman, J. C., Mook, W. G. & Vogel, J. C. in *Radiocarbon variations and absolute chronology* (ed. Olsson, I. U.) 275–301 (Wiley, New York, 1970).
2. Bacastow, R. & Keeling, C. D. in *Carbon and the biosphere* (eds Woodwell, G. M. & Pecan, E. V.) 86–135 (US Atomic Energy Commission Conference 720510, 1973).
3. Oeschger, H., Siegenthaler, U., Schotterer, U. & Guggenheim, A. *Tellus* **27**, 168–192 (1975).
4. Tans, P. P. & Mook, W. G. *Radiocarbon* **21**, 22–40 (1978).
5. Tans, P. P., de Jong, A. F. M. & Mook, W. G. *Nature* **271**, 234–235 (1978).
6. Damon, P. E., Long, A. & Wallick, E. I. *Earth planet. Sci. Lett.* **20**, 300–306 (1973).
7. Keeling, C. D. *Tellus* **25**, 174–197 (1973).
8. Lingenfelter, R. E. & Ramaty, R. in *Radiocarbon variations and absolute chronology* (ed. Olsson, I. U.) 513–537 (Wiley, New York, 1970).
9. Forbush, S. E. *Phys. Rev.* **70**, 771 (1946).
10. Forbush, S. E., Stinchcomb, T. B. & Schein, M. *Phys. Rev.* **79**, 501 (1950).
11. Münich, K. O. *Naturwissenschaften* **50**, 218–221 (1963).
12. Östlund, H. G. & Engstrand, L. G. *Radiocarbon* **5**, 203–227 (1963).
13. Nydal, R., Lövsæth, K. & Gulliksen, S. in *Proc. 9th int. Conf. Radiocarbon Dating*, California (in the press).

Life in the calcium chloride environment of Don Juan Pond, Antarctica

DON JUAN POND, which contains saturated calcium chloride brine, is in the south fork of the dry Wright Valley of Antarctica at latitude 77°33' S and longitude 167°10' E, and has been controversial almost since its discovery in 1961. Meyer *et al.*^{1,2} reported a sparse microflora of four species of heterotrophic bacteria and a yeast. Cameron, Horowitz and colleagues^{3,4}, using the Antarctic dry valleys as the best available natural simulation of Mars, reported that many areas were virtually sterile and, at most, limited to sparse bacterial populations. Field work^{5–10} has revealed a more abundant and varied microflora of yeasts, blue-green algae, fungi and bacteria, especially in the bottoms of frozen freshwater lakes. However, reports point to an extreme dry valley—exposed rock South-Polar biome consisting predominantly of heterotrophic forms, mainly prokaryotic with occasional fungal associates. If this is correct Don Juan Pond must, like the dry valleys generally, consist only of converter–consumer populations lacking extensive capability for continuous carbon reduction. Cameron has emphasised the ecological restrictions on the activity and distribution of algae in the dry valleys and mentions no algal or other autotrophic forms in his discussion of Don Juan Pond, even though thin organic layers were found 2 m below the 15–20 cm of standing water¹¹. However, in the course of a mercury sampling programme during the austral summer of 1978–79 we observed an extensive, irregular pellicle or mat-like structure 2–5 mm thick but extending 500–600 m² over much of the western part of the Don Juan Pond salt flats.

The extensive algal mat contained much adhering rock grains and fines and varied from yellow- to green- to black-brown. It could be lifted from the underlying sediment (Fig. 1, top). This mat was calculated to weigh about 1 tonne (0.167 g cm⁻¹) and contained about 130 kg of organic matter. Near the mat were moist areas of active gas bubble formation (Fig. 1, middle). Freshly moistened samples of dried mat contained ATP¹², proteins, polysaccharides and lipids, generally enclosing mineral particles. These samples also gave positive catalase and peroxidase reactions within seconds of moistening with appropriate reagents, and positive dehydrogenase within 45–60 min. In addition, dry mat placed on 1.5% agar containing 1% starch showed zones of hydrolysis when developed with I₂.KI after 4 h at 4°C. These zones increased with longer incubation and are indicative of β -amylase. Enzyme reactions^{13–16} of approximately equal intensity appeared whether mat samples were moistened with tap water, Don Juan Pond water or pure solutions of CaCl₂ and NaCl.

Microscopic examination of small flakes from the mat moistened with water or saline revealed densely entangled filamentous blue-green algae, principally an *Oscillatoria*-like organism (Fig. 1, bottom left), with non-mottled small blue-green filaments, non-motile green *Chlorella*-like cells, motile *Dunaliella*-like flagellates, non-motile golden-yellow 'tetrads' (Fig. 1, bottom right) and occasional phycomycetous hyphae and diatoms. Bacteria, both motile and non-motile, colourless and red-pigmented, were present but not abundant. At 24°C within 24 h after moistening flakes with distilled water, occasional Ciliata and Tardigrada were seen. Their absence in mat samples moistened with salt solutions may reflect delayed rather than inhibited development from the inactive state. After 5 d at 24°C, irrespective of salinity, most algal filaments and cells appeared faded, fragmented or lysed. Salt-free and low-salinity cultures were also overgrown with phycomycetes. At 4°C growth was slow but evident and cells maintained their osmotic integrity and pigmentation in Don Juan Pond water even after 2 weeks. The clear waters of Don Juan Pond seem naturally devoid of algae and fungi and have few bacteria, but nevertheless they contain significant amounts of ATP (unpublished results of P.L. and B.Z.S.)

The obvious explanation for the nearly barren condition of Don Juan Pond brine—that it is a consequence of osmotic stress combined with electrolyte imbalance—is not compelling. Many enzymes exhibit substantial activity in high salt media and

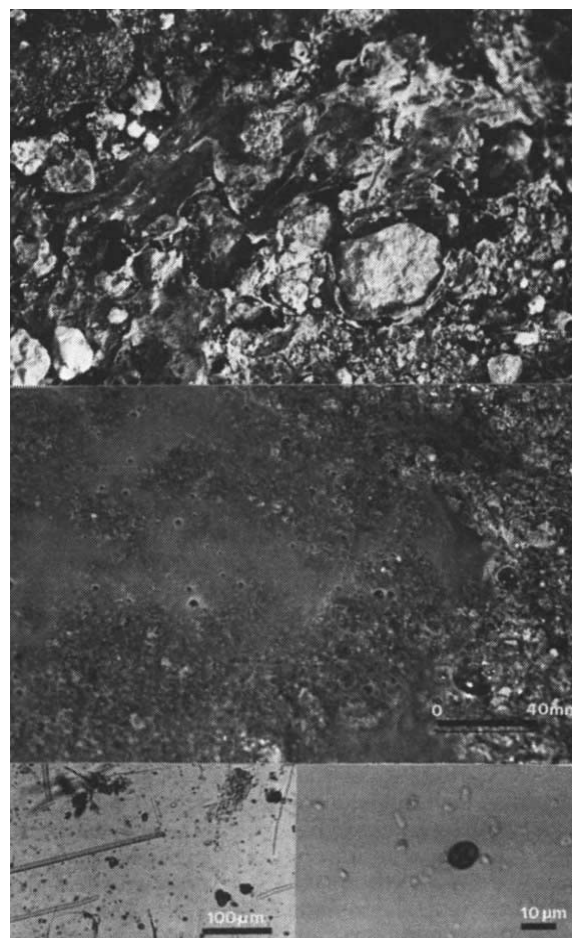


Fig. 1 Don Juan Pond and its microflora: top, surface view of the dry area showing pieces of mat around pebbles and rock fragments; middle, surface view of wet area showing bubble formation under free-standing brine; bottom left, *Oscillatoria*-like filaments and other algal forms in the mat; bottom right, golden-brown 'tetrads'.

Table 1 Major chemical constituents of the free liquid phase of Don Juan Pond

Constituent	Previous analysis (Meyer <i>et al.</i> ²)	New analyses	
		A	B
Ca ²⁺	114,000	167,500	159,000
Mg ²⁺	1,200	2,500	2,480
K ⁺	160	170	201
Na ⁺	11,500	7,050	5,580
Fe ³⁺	24	5.6	—
Mn ²⁺	0.05	2.7	—
Cu ²⁺	—	1.5	—
Co ²⁺	—	6.8	—
Ni ²⁺	—	9.2	—
Hg ²⁺	—	(3.5)	—
Cl ⁻	212,000	—	317,600
SO ₄ ⁻	11	—	—
NO ₃ ⁻	13	—	—
HCO ₃ ⁻	49	—	—
SiO ₂	—	—	Trace

All recent analyses were made by autoanalyser and atomic absorption spectrophotometry. Results are given in p.p.m. except for Hg which is in p.p.b. No effort was made to duplicate the sample site of Meyer *et al.* Analysis A was made at Hawaii Agricultural Experiment Station and analysis B at Hawaii Institute of Geophysics.

various microorganisms have been cultured in brines comparable with that from Don Juan Pond¹⁷⁻¹⁹. The minor constituents of Don Juan Pond—copper, cobalt and nickel—are in concentrations (1.5, 6.8 and 9.2 p.p.m., respectively) readily equal to if not greater than those required for toxic reactions in various plants and animals²⁰. The level of mercury is also relatively high (3.5 p.p.b.).

The gross average composition per 100 g of fresh mat was: 17.4 ± 7.4% water; 67.1 ± 7.2% mineral; 5.7 ± 1.8% soluble material, and 8.7 ± 2.5 organic matter ('loss on ignition'). Salts in the mat represent 32.7% dissolved solid material, or 327,000 p.p.m., compared with about 485,000 p.p.m. in the pond (Table 1, analysis B). Of the total soluble material, approximately 33% was precipitated as calcium carbonate which is the expected proportion for a saturated CaCl₂ solution. Spectrophotometry of samples of mat revealed major chlorophyll *a* bands at 410, 430, 475 and 663 nm, and minor bands at 580 and 615—together with accessory pigments or extraneous colouring matter. Paper chromatography (ether solvent) revealed up to six bands including yellow-brown and red-carotenoid-like fractions with double maxima around 500 nm. Using the putative organic fraction of the mat as the basis²¹, we calculated: in a dry sample, 105 and 1.4 µg per g of chlorophyll *a* and *b* respectively, and in a moist cold-preserved specimen, 265 and 4.9 µg per g, respectively.

Compared with typical chlorophyll values attained using *Oscillatoria* as a control—about 3,000 µg per g of cells—these are quite low. Nevertheless, preliminary manometric experiments have shown that the mat, in its natural medium at 4°C, produces oxygen in light (117 µE·cm⁻² s⁻¹ incandescent) and releases CO₂ in darkness.

These observations, although limited, provide evidence that the Don Juan Pond salt flats and brines have a relatively rich and trophically balanced ecosystem worthy of more intensive field and laboratory study.

We thank D. Thomas and R. Suehisa for analysing the brine. Figure 1, top and middle, were taken by PH2 E. Whitten, Naval Support Force Antarctica. This research was supported in part by an NSF-DPP grant (77-21507) to B.Z.S. and S.M.S.

B. Z. SIEGEL

Pacific Biomedical Research Center,
University of Hawaii,
Honolulu, Hawaii 96822

G. McMURTY

Hawaii Institute of Geophysics,
University of Hawaii

S. M. SIEGEL
J. CHEN

Department of Botany,
University of Hawaii

P. LAROCK

Department of Oceanography,
Florida State University,
Tallahassee, Florida 32306

Received 23 April; accepted 11 July 1979.

1. Meyer, G. H. *Polar Rec.* **11**, 317 (1962).
2. Meyer, G. H., Morrow, B., Wyss, O., Berg, T. E. & Littlepage, J. L. *Science* **138**, 110 (1962).
3. Cameron, R. E., Honour, R. C. & Morelli, F. A. in *Extreme Environments Mechanisms of Microbial Adaptation* (ed. Heinrich, M. R.), 57 (Academic, New York, 1976).
4. Horowitz, N. H., Cameron, R. E. & Hubbard, J. S. *Science* **176**, 242 (1972).
5. Friedmann, E. I. *Antarctic J.* **12**, 26 (1977).
6. Friedmann, E. I. & Ocampo, R. *Science* **193**, 1247 (1976).
7. Holm-Hansen, O. *Phycologia* **4**, 42 (1964).
8. Vishniac, W. V. & Mainzer, S. E. *Antarctic J.* **7**, 88 (1972).
9. Drouet, F. *Polar Rec.* **11**, 320 (1962).
10. Uydess, I. L. & Vishniac, W. V. in *Extreme Environments, Mechanisms of Microbial Adaptations* (ed. Heinrich, M. R.), 29 (Academic, New York, 1976).
11. Mudrey, M. G., Jr, Torii, T. & Harris, H. *Dry Valley Drilling Project (DVDP) Bulletin*, 85 (August, 1975).
12. Karl, D. M. & LaRock, P. A. *J. Fish. Res. Bd Can.* **32**, 599 (1975).
13. Casselman, W. G. *Histochemical Techniques* (Methuen, London, 1959).
14. Glick, D. *Quantitative Chemical Techniques of Histo- and Cytochemistry* (Interscience, New York, 1963).
15. Hawk, P. B. *Physiological Chemistry*, 4th edn. (Blakiston, Philadelphia, 1914).
16. Pearse, A. G. *Histochemistry*, (Churchill, London, 1968).
17. Siegel, S. M. in *The General and Comparative Biology of Terrestrial Organisms under Experimental Stress Conditions*, Semiannual Report Contract No. NASw-767, National Aeronautics and Space Administration, 1 (1965).
18. Siegel, S. M., Speitel, T. & Stoeker, R. *Cryobiology* **6**, 160 (1969).
19. Siegel, S. M. *Univ. Hawaii Botanical Science Paper No.* 31, 1 (1973).
20. *Handbook of Environmental Control 3 Water Supply and Treatment* (ed. Bond, R. & Straub, C.) 345-388 (CRC Press, Cleveland, 1973).
21. *Biochemist's Handbook* (ed. Long, C.) 1029-1032 (Van Nostrand, Princeton, 1961).

Texture contours can facilitate stereopsis by initiating vergence eye movements

IN man the two eyes receive slightly differing views of the world. The visual system utilises this disparity for stereoscopic depth perception. This involves a difficult correspondence problem for establishing binocular fusion: for each surface feature extracted from one eye's image, the brain has to decide which is the equivalent feature in the other eye's image. Ambiguities often exist, particularly for densely-textured visual surfaces, but one aspect of binocular visual processing which helps reduce the scale of the problem is that for any given vergence angle, the visual system considers only those potential left/right fusions with disparities under a certain limiting size (defined by Panum's fusional areas^{1,2}). This disparity limit, however, requires that when disparities greater than Panum's area are presented, a suitable vergence change must be initiated to bring corresponding features within the allowable range for fusion. A question currently of considerable theoretical interest concerns the nature of the stimulus characteristics used to initiate these essential vergence movements. The interest stems mainly from Marr and Poggio's model of human stereo vision^{3,4} in which false correspondences are avoided by seeking left/right fusions within spatial frequency tuned disparity channels whose disparity range is closely tied to their spatial frequency sensitivity⁵. By setting the disparity range of any given channel to about $\pm w$ (where w = the width of the centre of the receptive fields of the channel in question) Marr and Poggio have demonstrated that false matches are improbable and any that do emerge are easily eliminated using quite simple rules. A further feature of the model is that the outputs of low spatial frequency tuned channels (that is units capable of dealing with large disparities because of

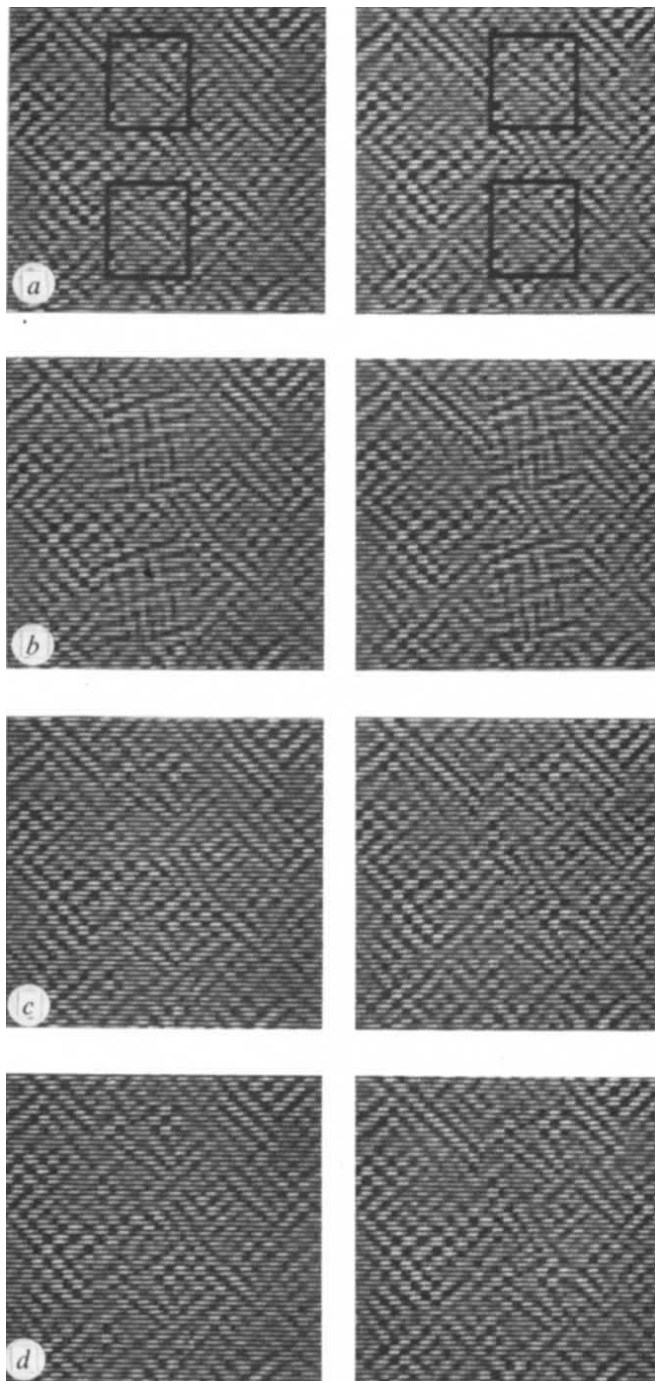


Fig. 1 Photographs of the oscilloscope displays showing examples of stereograms used in the two experiments as follows: *a*, Line contour stereogram. *b*, Texture contour stereogram. *c*, Cyclopean stereogram. *d*, Cyclopean stereogram used for the depth discrimination task in the second experiment. All stimuli were computer-generated in two steps. First, oriented textures were produced by filtering random-dot patterns using gaussian filter functions tuned for orientation and spatial frequency (centre frequency 4.5 cycles deg^{-1} , obtained here by viewing the illustrations from about $11 \times$ picture height; bandwidth ± 0.9 cycles deg^{-1} ; orientation bandwidth $\pm 10^\circ$). Textures either had component orientations of 45° and 135° (for example *c* surround) or 75° and 165° (for example, *c* targets). This filtering was done using a 128×128 pixel array for each half. The second step in stimulus production was to reduce each stereo half to a 64×64 pixel array (the resolution appearing in the illustrations), using a sampling procedure which had no important effects on spatial frequency content as judged by examining the resulting power spectra and by obtaining equivalent psychophysical data in pilot trials run using a tachistoscope to present the original 128×128 pixels stimuli as hardcopy outputs produced by a high quality facsimile receiver⁶. The reduction to 64×64 pixels permitted on-line stimulus presentation with a satisfactory refresh rate (20 ms per frame). Stereograms *a*, *b* and *c* contained upper and lower targets, each with a mean disparity of ~ 35 arc min (the target elements lay in the disparity range 32.7 to 37.4 arc min as a side effect of the sampling procedure used to obtain the 64×64 pixel arrays). Stereogram *d* contained an upper target with a disparity of about 32.7 arc min and a lower target with a disparity of about 37.4 arc min, so presenting the depth discrimination task described in the text. In all stereograms apart from *a*, disparate targets were inset into their surrounds using a cosine contrast ramp over a period of 6 pixels (within the 128×128 array) to preclude discrimination of a monocular contour in the texture contour condition on the basis of trivial edge effects at the boundary of the two textures.

In our first experiment, we used a nonius line technique for recording vergence changes made in response to tachistoscopic presentations of various stereograms. The stimuli were generated by computer on-line and presented via a pair of oscilloscope displays (one display for each 5×5 cm stereo half) which were viewed at about 57.5 cm and binocularly superimposed using Risely prisms. A trial began with the subject fixating a central fixation spot; above this was a black nonius line (20 mm long, 0.8 mm thick) seen only by the right eye and below was an identical line seen only by the left eye. When ready to proceed (having adopted suitable fixation, as demonstrated by the nonius lines appearing vertically aligned), the subject pressed a hand-held switch. He then heard a 1-s warning tone which was followed after a 2-s pause by an 80-ms presentation of one of the stereograms shown in Fig. 1*a–c*. This presentation was followed by a 40 ms uniform grey field of the same space average luminance as the stereograms (about 4.3 cd m^{-2}). Finally, a second pair of nonius lines with no central fixation spot was displayed for a further 400 ms. The subject's task was to estimate the direction and extent (in mm) of any apparent misalignment of the second pair of nonius lines at their first moment of appearance. (Any subsequent movement of the nonius lines was ignored; further experiments⁶ using 60-ms nonius line exposures to remove the problem of nonius shifts during the 400-ms exposure period have confirmed the findings we report here.) A convergent eye movement was indicated if the top line appeared misaligned to the left, vice versa for a divergent movement.

The stereograms contained two disparate targets, one positioned in the upper half field and one in the lower half field (two targets were used for comparison with stimuli used in the second experiment). Both upper and lower targets in each stereogram had the same disparity (either $\sim 35'$ convergent or $\sim 35'$ divergent), except for control conditions in which both targets had zero disparity. Targets were either cyclopean (not visible until after binocular fusion, as in Fig. 1*c*) or outlined by either a line or texture contour (Fig. 1*a* and 1*b* respectively). The different contour and disparity conditions were randomly presented, with each subject having a total of eight trials on each condition. Data were collected from two experienced observers and Fig. 2*a* shows the mean size and direction of misalignment of the nonius lines following the various stereogram presentations. For both subjects, disparate texture and line contours initiated an appropriate vergence eye movement (that is, one which would lead to bifoveal fixation of the targets). In contrast, the zero disparity texture and line contours produced little if any misalignment of

their relatively large receptive fields) can initiate vergence movements bringing into correspondence higher spatial frequency channels mediating smaller disparities. The theory predicts that in the case of a two-planar densely-textured stereogram (for example Fig. 1*a–c*) in which the disparity range exceeds the size appropriate for the lowest spatial frequency channels activated by the stereogram's texture, vergence movements should exhibit only a random-search structure. We report here an experimental test of this prediction which shows that vergence movements are not always random in the stimulus circumstances just specified. Rather, vergence can be guided by texture contours (Fig. 1*b*) depicted by differences of texture orientation between regions of relatively high spatial frequency content for the disparity range incorporated in the stereogram. Moreover, we find that such texture contours can facilitate stereopsis in a way consistent with the idea that they do so by initiating appropriate eye movements.

the nonius lines. Also, all three cyclopean conditions—convergent, divergent and zero disparity—failed to create much misalignment of the nonius lines. Note that nonius alignment/misalignment appeared involuntarily for both subjects. Also, as the task of judging the size of the various shifts required the subjects' full concentration, they were unable to notice any depth effects. However, additional trials omitting the post-stereogram nonius lines have demonstrated that both subjects can reliably discriminate between convergent and divergent disparities from 80-ms exposures of the line and texture contour stereograms but not from similar exposures of the cyclopean stimuli.

The fact that no vergence movements of any importance were initiated by the flashed cyclopean stereograms is in accord with Marr and Poggio's theory. For two-planar disparities of the size we used (approximately $\pm 35'$), the theory predicts vergence movements to be triggered if a disparity channel tuned to spatial frequencies under 1 cycle deg^{-1} was activated. The texture for our stereograms was created using narrow-band filters tuned to $4.5 \text{ cycles deg}^{-1}$, well above the required range, and so the absence of vergence eye movements for the cyclopean conditions is in keeping with the theory's prediction. On the other hand, flashed disparate texture contours did initiate appropriate vergence movements, even though these texture contours did not alter the spatial frequency range of the stimulus texture, implying that human stereopsis does not rely solely on low spatial frequency/large disparity mechanisms to guide vergence eye movements.

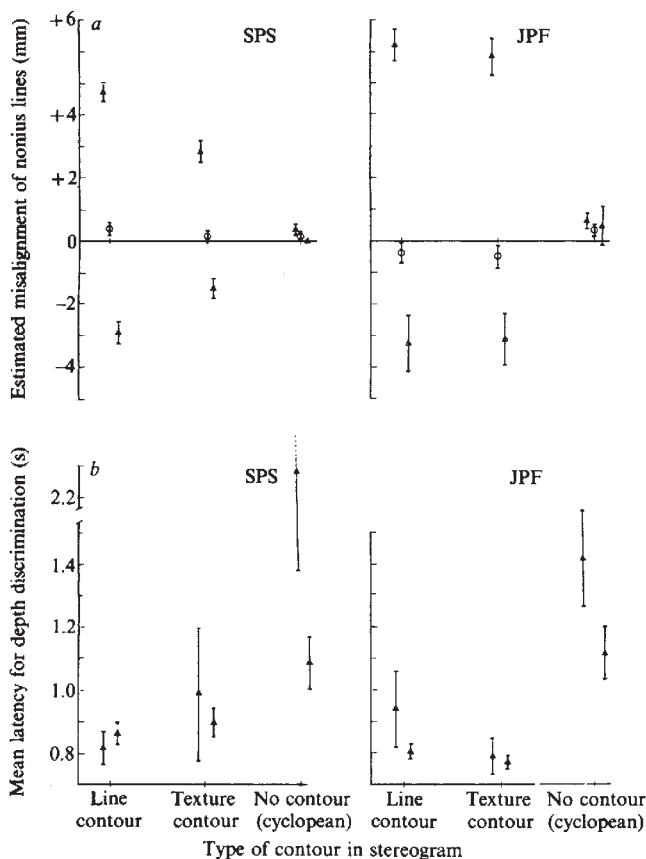


Fig. 2 *a*, Estimated misalignment of the nonius lines following 80-ms presentations of the stereograms shown in Figs 1*a-c*. ▲, convergent disparity; △, divergent disparity; ○, zero disparity. The bars show one standard error to either side of the mean. *b*, Reduction in mean latencies for a cyclopean depth discrimination task (Fig. 1*d*) following 80-ms presentations of stereograms in Fig. 1*a-c* (▲, convergent disparity; △, divergent disparity; ○, zero disparity). bars = ± 1 s.e.m.

Given the above results, it becomes of interest to discover whether a vergence movement triggered by a tachistoscopic presentation of a line or texture stereogram can facilitate a subsequent cyclopean depth discrimination task which requires this vergence movement for its successful execution. Consequently, in our second experiment we measured latencies to make a cyclopean depth discrimination following tachistoscopic presentations of the stereograms used in the first experiment. Experiments began with the subject fixating a central spot. When he was ready to proceed, he pressed a hand-held switch which initiated after a 2-s pause, an 80-ms presentation of one of the stereograms shown in Fig. 1*a-c*. This stimulus was followed by a 40-ms presentation of a uniform grey field of same space average luminance as the stereograms (as before, about 4.3 cd m^{-2}), after which came a cyclopean stereogram presenting a depth discrimination task (Fig. 1*d*). In this latter stereogram, both targets possessed either a convergent or divergent disparity, as in the flashed disparate stereograms that preceded them, but now the targets differed slightly in depth due to a disparity difference of about $4.7'$. This difference was created by setting the upper target to a disparity of $\sim 37.4'$ and the lower target to a disparity of $\sim 32.7'$, or vice versa. On the appearance of this cyclopean discrimination stimulus, the subject's forced-choice task was to press a switch up or down to indicate whether the upper or lower target appeared the nearer. This response enabled his stereopsis latency to be measured and it was also used to terminate the cyclopean presentation. Note that the design of the experiment was such that on half the trials a convergent up/down discrimination was required and on the remaining trials a divergent one. In every case, however, the subject's task was to press the switch up or down to indicate which target was the nearer. As in the previous experiment, the different contour and disparity conditions were presented randomly, with each subject having eight trials on each condition.

Data for the same two practised observers are shown in Fig. 2*b*. In conditions where the flashed stereogram contained a line or texture contour outlining the disparate targets, the mean latency for stereopsis was significantly reduced in comparison with the control condition where only cyclopean targets were flashed (for both subjects, $F > 12$, $d.f. = 2, 42$, $P < 0.0001$). Comparison of the data from our two experiments thus strongly suggests that the facilitation of stereopsis evident in Fig. 2*b* is a result of the flashed line and texture contours initiating vergence eye movements as recorded in Fig. 2*a*, movements which would be appropriate for the fusion of the immediately-following cyclopean stereogram. These data thus extend into a tachistoscopic paradigm the already known capability of continuously-present monocular features for facilitating stereopsis from random-dot stereograms^{7,8}.

For simplicity, we have described our second experiment as though only appropriate disparate flashed cues were used. However, the experiment also included zero disparity and wrong disparity cues (in the latter case, a divergent cue preceded a convergent cyclopean presentation, or vice versa). As would be expected from the vergence hypothesis under consideration, these cues provided either no facilitation or even a hindrance to successful fusion (full data will be reported elsewhere⁶).

A further condition in our second experiment showed that flashed texture cues carried by rivalrous texture targets composed of differently-orientated elements in the two fields⁶ provided just as good facilitation as the same-orientation targets shown in Fig. 1*c*. Thus S.P.S. produced mean latencies (seconds) for the different- and same-orientation target conditions of 0.95 (s.e.m. = 0.10) and 0.86 (0.05) respectively (convergent and divergent data pooled). Equivalent means for J.P.F. were 0.82 (0.74) and 0.78 (0.03). This result suggests that some instances of qualitative stereopsis from rivalrous texture stereograms⁹⁻¹² may well be mediated by mechanisms whose primary function is the control of vergence shifts.

We conclude that disparate texture cues can trigger vergence movements appropriately and that their ability to facilitate

stereopsis is probably based at least partly on this fact. This conclusion implies that it is insufficient to limit vergence control to spatial frequency channels whose tuning determines their disparity selectivity. Rather, it seems that any reasonably complete model of human stereopsis must allow for the early computation of information about boundaries separating regions of different textures, and it must allow the vergence eye movement control mechanism rapid access to this information.

We thank Dr Christopher Brown for construction of the on-line computer displays and Philip Stenton for his help as a subject.

A. L. KIDD
J. P. FRISBY
J. E. W. MAYHEW

Department of Psychology,
University of Sheffield,
Sheffield, UK

Received 18 April; accepted 16 July 1979.

1. Ogle, K. N. *Archs Ophthalmol.* **48**, 50 (1952).
2. Fender, D. & Julesz, B. *J. opt. Soc. Am.* **57**, 819 (1967).
3. Marr, D. & Poggio, T. *Proc. R. Soc. B204*, 301 (1979).
4. Grimson, W. E. L. & Marr, D. *Proc. Image Understanding Workshop* (Inf. Proc. Tech. Office of DARPA, 1979).
5. Felton, T. B., Richards, W. & Smith, R. A. Jr. *J. Physiol., Lond.* **225**, 349 (1972).
6. Frisby, J. P., Mayhew, J. E. W. & Kidd, A. L. (in preparation).
7. Saye, A. & Frisby, J. P. *Perception* **4**, 159 (1975).
8. Julesz, B. & Oswald, H. P. *Perception* **7**, 315 (1979).
9. Ramachandran, V. S., Madhusudhan Rao V. & Vidyasagar, T. R. *Nature* **202**, 412 (1973).
10. Mayhew, J. E. W. & Frisby, J. P. *Nature* **264**, 53 (1976).
11. Frisby, J. P. & Mayhew, J. E. W. *Perception* **7**, (1978).
12. Kaufman, L. *Sight and Mind* (Oxford University Press, London, 1974).

The herring ear has a unique receptor pattern

HERRINGS and their relatives in the teleost order Clupeiformes, including sardines and anchovies, have an unusual connection between the gas bladder and the inner ears^{1,3} that may indicate an auditory mechanism involving the utricle, a part of the ear generally considered to be a gravistatic organ in all other vertebrates^{4,5}. In herring, two slim ducts extend forward from the gas bladder, one to each ear, and connect with gas-filled portions of the otic bullae within the skull. The membranous wall of the utricle is in contact with these pro-otic and pterotic bullae, and thus might be stimulated directly by vibrations of this compressible system. Structural features of the ducts and bullae in clupeids suggest that certain acoustic frequencies may be selectively filtered^{6,7}, but this requires electrophysiological confirmation⁸. However, although such studies^{1,6,7} support the argument of an auditory function for the clupeid utricle, no data have been available regarding ultrastructural features of the herring utricle. We now report that scanning electron microscopic (SEM) investigations of the otolith sensory epithelial surfaces have revealed several additional structural aspects of the herring utricle which may further support a hypothesis for an auditory role for this organ in clupeids. These features seem to be unique to the clupeid utricle and include the presence of large numbers of sensory cells with long ciliary bundles, and a special sensory receptor pattern.

Inner ear sensory surfaces from the Pacific herring, *Clupea harengus pallasii* (a subspecies of the worldwide *C. harengus*), were prepared as in other studies^{9,10}. Tissues were fixed in buffered glutaraldehyde solution, post-fixed in buffered osmium tetroxide, and after removal of the statoconia (otoliths) the tissues were dehydrated, then dried by the critical-point method using carbon dioxide. Observation, photography and mapping were done with the scanning electron microscope. Additional data from other clupeiform representative species¹¹ fully agree with our data for herring.

An unusual feature of the herring utricle is the partition of the sensory epithelium into three laterally directed 'fingers' which



Fig. 1 Tall ciliary bundles typical of those found on herring utricular sensory epithelia ($\times 8,000$).

are separately innervated². These three regions, named the macula anterior, macula media and macula posterior², are shown by SEM observations to be three distinctly separate populations of ciliated mechanosensory hair cells. Even on the medial wall from which the fingers diverge there is a gap without sensory cells between each of these three maculae.

A second unusual feature in the herring utricle is a widespread distribution of a relatively 'tall' form of ciliary bundle. Each bundle, as in other vertebrate hair cells, contains a single kinocilium which is a true cilium, located at one end of an orderly array of stereocilia which are similar to microvilli; the stereociliary array consists of rows of stereocilia which decrease in height the further they are from the kinocilium. In the herring utricle there are large numbers of bundles which have a kinocilium at least 6–7 μm long and some stereocilia which are nearly as long (Fig. 1). There is a relatively dense forest of these tall bundles over much of the macula media, and they are also common on the maculae anterior and posterior. Near the edges of all three maculae, but more broadly inward from the margins of the maculae anterior and posterior, there is another form of bundle with a still longer kinocilium but short (1–2 μm) stereocilia.

The pattern of hair cell orientations in the herring utricle also seems to be a feature unique to this group. The asymmetry of the ciliary bundle allows determination of a morphological polarisation vector for each hair cell¹². In each of the three maculae in the herring utricle the cells can be divided into two populations of opposing orientations (Fig. 2). For example, hair cells in the anterior half of the macula media are orientated in a generally posterior direction (with respect to the axis of the fish), whereas those in the posterior half are orientated anteriorly. An imaginary, nearly straight line can be drawn to divide the opposing populations, running in a medial to lateral direction, and bisecting the macula media in the transverse axis of the fish. A similar but more curved line can be drawn for the anterior and posterior maculae, also transverse to the fish and following the curve of these maculae. Although the map in Fig. 2 shows a flat projection of the maculae, the utricular pouch is actually quite rounded. As the anterior macula lies well up the slope of the anterior utricular wall, its hair cells should be considered to have considerable upward or downward orientation vectors, rather than simply anterior-posterior vectors.

We now have sufficient comparative data to put the features of the herring utricle in some perspective. In all vertebrates studied to date the utricular macula is a single dish or crescentic patch, sometimes with a postero-lateral process, the lacinia⁴. Generally, there is a distinctive antero-lateral band, the striola, in which the hair cells differ markedly in size or in ciliary bundle form from the hair cells in the rest of the macula¹³. Further, the usual orientation map of the vertebrate utricle, in both bony

fishes^{9,12,14,15} and tetrapods^{13,16}, shows a wide antero-lateral area containing hair cells with their kinocilia on the side inward from the periphery, whereas the large medial-posterior area has hair cells orientated facing outward towards the periphery. The dividing line between these two populations curves along the middle of the striola (and lacinia, when present), usually in a lateral to posterior curve that parallels the macular edge. An exception to this pattern occurs in at least one elasmobranch fish, the thornback ray, which has a mixture of patches of opposing orientations in the utricle¹⁷, although the macula is not structurally split.

The presence of the tall ciliary bundles and the absence of a striola in the herring utricle also occur in the herring sacculus and lagena¹¹. In several other teleosts these features are found in the sacculus and lagena¹⁰, whereas the utricle retains the usual striola and lacks a large proportion of tall bundles^{9,14,15}. The functional significance of bundle morphology is unclear, although there is some evidence that tall bundles may be sensitive to a higher frequency range than shorter bundles¹⁸. However, bundle forms may depend for both their form and function on properties of the overlying otolith membrane.

The orientation patterns of the herring utricular maculae may have phylogenetic origins in the general vertebrate utricular pattern. This is suggested by the fact that all vertebrate utricles have cells orientated towards the line of division between populations, and that the herring macula posterior has a gross location and polarisation pattern similar to that of the lacinia in other fishes. The utricular opposition in herring differs significantly from that in the sacculus in clupeids¹¹ and many other bony fishes, where hair cells are orientated away from lines of population division^{9,10,14,15}.

Because the directional sensitivity to shear has a morphological basis in the asymmetry of the ciliary bundle, the orientation patterns in the vertebrate ear are thought to have a functional significance^{19,20}. The macula anterior of the herring utricle, with its opposed populations on the anterior wall, may be a bidirectional detector for dorso-ventral shearing displacements, and the large macula media could be the major bidirectional detector for rostral-caudal displacements. The sensitivity of this arrangement could be enhanced by the presence of the pro-otic fenestra from the pro-otic bulla that forms a fluid connection to the utricle⁷. This fenestra opens between the macula anterior

and macula media and is in a prime location for driving displacements of hair cell bundles on those maculae.

These ultrastructural findings indicate that hair cell ciliary bundles and hair cell orientation patterns in the herring utricle are distinctly different from the utricular structure of any other vertebrate yet examined. Coupled with the known relationship between the gas bladder, bullae and utricle, these features are not incompatible with the argument that the herring utricle may respond to acoustic stimuli. Sensitivity could be enhanced by the gas-filled otic bullae, and input processed by the complex differentiation of the tripartite macula in a manner unique among vertebrates.

This work was supported by grants from NINCDS (NS-15090) and NSF (BNS78-22411) to A.N.P., NINCDS (1 RO1 13946-02) to C.P., and a Research Career Development Award from NIH (NS-00312) to A.N.P.

ARTHUR N. POPPER

Department of Anatomy,
Georgetown University School of
Medicine and Dentistry,
3900 Reservoir Road NW,
Washington DC 20007

CHRISTOPHER PLATT

Department of Biological Sciences,
University of Southern California,
Los Angeles, California 90007

Received 23 April; accepted 5 July 1979.

- Allen, J., Blaxter, J. H. S. & Denton, E. J. *J. mar. biol. Ass. U.K.* **56**, 471-486 (1976).
- O'Connell, C. P. *U.S. Fish Wildl. Serv. Fish. Bull.* **56**, 505-533 (1955).
- Wohlfahrt, T. A. Z. *Morph. Ökol. Tiere* **31**, 371-410 (1936).
- Lowenstein, O. *Handbook of Sensory Physiology*, Vol. 6, Pt 1 (ed. Kornhuber, H. H.) 75-120 (Springer, Berlin, 1974).
- von Holst, E. Z. *vergl. Physiol.* **32**, 60-120 (1950).
- Denton, E. J., Gray, J. A. B. & Blaxter, J. H. S. *J. mar. biol. Ass. U.K.* **59**, 27-47 (1979).
- Denton, E. J. & Blaxter, J. H. S. *J. mar. biol. Ass. U.K.* **56**, 787-807 (1976).
- Enger, P. S. *Comp. Biochem. Physiol.* **22**, 527-538 (1967).
- Platt, C. J. *comp. Neurol.* **172**, 283-298 (1977).
- Popper, A. N. *J. Morph.* **153**, 397-417 (1977).
- Platt, C. & Popper, A. N. in *Vestibular Function and Morphology* (ed. Gualtierotti, T.) (Springer, Berlin, in the press).
- Flock, A. *J. Cell Biol.* **22**, 413-431 (1964).
- Lindeman, H. H. *Ergben. Anat. EntwGesch.* **42**, 1-113 (1969).
- Dale, T. *Norw. J. Zool.* **24**, 35-128 (1976).
- Popper, A. N. *J. comp. Neurol.* **181**, 117-128 (1978).
- Lewis, E. R. & Li, C. W. *Brain Res.* **83**, 35-50 (1975).
- Lowenstein, O., Osborne, M. P. & Wersäll, J. *Proc. R. Soc. Lond.* **B160**, 1-12 (1964).
- Weiss, T. F., Mulroy, M. J., Turner, R. G. & Pike, C. L. *Brain Res.* **115**, 71-90 (1976).
- Hudspeth, A. J. & Corey, D. P. *Proc. natn. Acad. Sci. U.S.A.* **74**, 2407-2411 (1977).
- Lowenstein, O. & Wersäll, J. *Nature* **184**, 1807-1808 (1959).

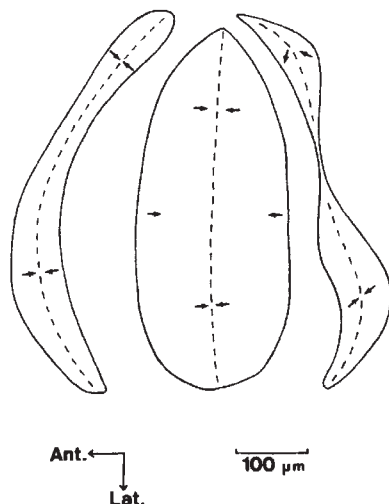


Fig. 2 Map of hair cell orientation patterns in the herring utricle (left ear, 50 mm fish). Arrows correspond to the morphological polarisation of hair cell populations in a sampled region, given by the asymmetric location of the kinocilium on the hair cell ciliary bundles. Dashed lines indicate divisions between populations of different orientations; two opposed populations occur in each of the three maculae. Note that this is a schematic flat projection map of surfaces from the rounded utricular pouch; the macula anterior actually lies on the curve of the anterior wall, giving its cells a substantial dorso-ventral orientation component.

A phosphoglycolate phosphatase-deficient mutant of *Arabidopsis*

CURRENT concepts of photosynthetic and photorespiratory carbon metabolism have been derived almost entirely from the results of biochemical and physiological investigations. To examine these concepts by genetic analysis, we have isolated several mutants of a C_3 species with defects in CO_2 assimilation and photorespiration. We report here the isolation and preliminary characterisation of the first of these, a mutant deficient in phosphoglycolate phosphatase activity.

The mutants were recovered as conditional lethals which are phenotypically indistinguishable from the wild type when grown in air enriched to 1% CO_2 , but which are inviable in normal atmospheric conditions of 0.03% CO_2 . The rationale for the screening procedure is that conditions of elevated CO_2 limit ribulosebiphosphate (RuBP) oxygenase activity¹, thereby reducing the flow of carbon through the photosynthetic carbon oxidation (PCO) cycle². As chemical inhibition of enzymes in the PCO cycle reduces the photosynthetic rate in conditions which normally permit photorespiration, but does not affect photosynthesis at high CO_2 or low O_2 concentrations³, we postulated that mutants with defects in the PCO cycle would be

viable only in conditions of high CO_2 or low O_2 . These growth conditions should also permit the recovery of mutants with a greatly increased ratio of RuBP oxygenase to RuBP carboxylase activity.

Wild-type *Arabidopsis thaliana* (L.) Heynh. seed was mutagenised with ethyl methane sulphonate⁴, brought to maturity in normal cultural conditions, and bulk collected. The M2 generation was grown to the four-leaf stage in air enriched to 1% CO_2 . At this time, chlorotic plants and those which lacked vigour were removed. The CO_2 concentration was reduced to 0.03%, and plants which remained normal in appearance after 4 d were discarded. The remaining plants were returned to a 1% CO_2 regime for 4 d. Those plants which failed to recover

Table 1 Products of $^{14}\text{CO}_2$ assimilation by wild-type and mutant *Arabidopsis*

	Wild-type	CS119	Wild-type HBA-treated*	CS119 HBA- treated*
Basic fraction	21.4†	11.0	3.2	7.6
Glycine	10.5	0.5	0.3	0.4
Serine	8.4	0.9	0.3	0.5
Alanine	2.0	9.0	1.4	5.9
Neutral fraction	6.1	14.8	3.6	9.3
Acid-1 fraction	36.7	23.3	61.7	23.9
Glycolate	2.7	0.9	28.8	1.8
Acid-2 fraction	8.8	29.1	6.2	28.1
Phosphoglycerate	6.3	5.6	3.8	7.8
Phosphoglycolate	0.7	19.6	0.7	19.6
Acid-3 fraction	24.4	17.6	23.7	27.4
Recovery	97.4	95.8	98.4	96.3

Intact plants, grown as described in Fig. 1, were removed to a photosynthesis chamber and flushed with $400\ \mu\text{l l}^{-1}$ CO_2 , 21% O_2 and bal. N_2 at 25°C , 75% relative humidity (RH), at $300\ \mu\text{einsteins m}^{-2}\text{s}^{-1}$ for 2 min. At this time the system was closed and $^{14}\text{CO}_2$ introduced. After 2 min of $^{14}\text{CO}_2$ incorporation, the plants were quickly (3–5 s) removed to liquid nitrogen, then extracted by grinding in 4 M formic acid. The CO_2 concentration in the chamber remained at $350\text{--}400\ \mu\text{l l}^{-1}$ during the period of incorporation. The extract was lyophilised and the residual water-soluble compounds were fractionated by ion exchange and paper chromatography as described previously³ except that the acid fractions were recovered from a Dowex-1 (formate) column by successive elution with 10 ml 4 M formic acid (acid-1 fraction), 10 ml 8 M formic acid (acid-2 fraction) and 6 ml 4 M HCl (acid-3 fraction). Phosphoglycolate was identified by co-migration with authentic samples in several solvent systems, before and after phosphate ester hydrolysis.

* Inhibition of *in vivo* glycolate oxidase activity was accomplished by submerging the roots of intact plants in 5 mM 2-hydroxy-3-butyroic acid (HBA)¹⁶ for 20 min in the elevated CO_2 regime described in Fig. 1. Following treatment, the plants were labelled, extracted and fractionated as above. Control experiments, in which identically treated wild-type and mutant plants were assayed for glycolate oxidase activity, indicated that the HBA treatment resulted in >95% inhibition of enzyme activity.

† Values indicate % of total water-soluble counts recovered, and represent the mean of two independent experiments.

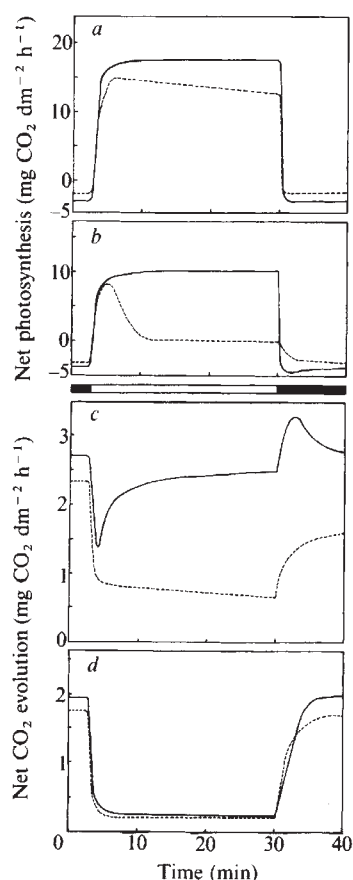


Fig. 1 Net CO_2 exchange of wild-type and mutant *Arabidopsis* in various gas regimes. Plants were grown in 1% CO_2 , 21% O_2 and bal. N_2 at 25°C and 75% RH under continuous fluorescent illumination ($170\ \mu\text{einsteins m}^{-2}\text{s}^{-1}$) on vermiculite irrigated with half-strength Hoagland's solution. Gas exchange was monitored on intact plants (at floral initiation stage) from which the lower leaves had been removed. The plants were placed in a glass chamber of 20 ml volume and continually flushed with a humidified (75% RH) gas stream at a flow rate ($100\text{--}150\ \text{ml min}^{-1}$) sufficient to maintain a ΔCO_2 of less than $15\ \mu\text{l l}^{-1}$. The CO_2 concentration of the exiting gas was continuously monitored with an IR gas analyser. The total volume of the system, including measuring cuvette, was 46 ml. Light ($300\ \mu\text{einsteins m}^{-2}\text{s}^{-1}$) was provided by a 300 W tungsten lamp filtered through 15 cm of water. The plant chamber was maintained at 25°C by immersion in circulating water. The composition of gas entering the chamber was: a, $355\ \mu\text{l l}^{-1}$ CO_2 , 2.0% O_2 , bal. N_2 ; b, $352\ \mu\text{l l}^{-1}$ CO_2 , 21% O_2 , bal. N_2 ; c, 50% O_2 , bal. N_2 ; d, 2% O_2 , bal. N_2 . The response of the wild-type plants is indicated by the solid line, and that of the mutant by the broken line. The open or closed bar inserted in the figure represents light or dark conditions, respectively.

were discarded. This procedure resulted in a 500-fold reduction in population size. The remaining plants were taken to maturity in 1% CO_2 without further screening. A similar screening procedure, applied in the M3 generation to the individual lines established from the M2, resulted in the retention of 40 probable mutants from an initial population of approximately 40,000 M2 plants. The putative mutants were advanced to the M5 generation by single seed selection without additional screening.

A preliminary survey of the gas exchange characteristics of the M5–M7 generations revealed that most of the putative mutants exhibited altered photosynthesis characteristics. The mutant strain CS119 was particularly striking in this respect. In a gas regime of $355\ \mu\text{l l}^{-1}$ CO_2 , 2.0% O_2 and balance N_2 , the mutant assimilated CO_2 for extended periods at a rate about equal to that of the wild type (Fig. 1a). In an atmosphere containing 21% O_2 , the photosynthesis rate of the mutant initially approached the wild-type rate but then rapidly declined to zero (Fig. 1b). This reduction of photosynthetic capacity was completely reversed by a 15–20-min dark treatment imposed at the time of 50–70% inhibition (data not shown). The rapidity of onset and the reversibility of inhibition suggest that the response is caused by a reversibly altered metabolite pool rather than photo-oxidative damage. Although we have not measured photosynthesis in 1% CO_2 , normal growth of the mutant in such conditions implies that high levels of CO_2 can alleviate the deleterious effects of O_2 on photosynthesis. Thus, the mutant is

capable of sustained photosynthesis only in conditions which limit RuBP oxygenase activity^{1,5}.

The mutant also exhibited a greatly reduced rate of CO₂ release during illumination and did not show the post-illumination burst of CO₂ release commonly ascribed to photorespiration⁵ (Fig. 1c). The rate of CO₂ release by the mutant in 2% O₂ (Fig. 1d) is about 30% of that observed in 50% O₂, suggesting that a substantial proportion of residual respiration in light is due to incomplete suppression of dark respiration during illumination⁶. Compared with other species⁵, both wild-type and mutant plants exhibited a disproportionately high rate of CO₂ evolution in the dark. We have tentatively ascribed the relatively high dark respiration rates to the 24-h photoperiod used in cultivation of the plants.

The probable site of the lesion in the mutant was deduced from the pattern of ¹⁴CO₂ incorporation into early products of photosynthesis (Table 1). After 2 min of labelling in normal atmospheric conditions, 20% of the ¹⁴C incorporated by the mutant was found in phosphoglycolate. In contrast, phosphoglycolate was barely detectable in wild-type *Arabidopsis*, as it is in other species examined under similar circumstances^{3,7}. The increased labelling of the phosphoglycolate pool was accompanied by a large decrease in the labelling of serine and glycine, products of photorespiratory glycolate metabolism^{2,8}. These results suggested that the mutant was deficient in phosphoglycolate phosphatase, a chloroplast enzyme which specifically hydrolyses phosphoglycolate to glycolate⁹. This conclusion was confirmed by assaying phosphoglycolate phosphatase activity in crude S-30 leaf extracts. The activity in extracts of the mutant was less than 5% of that found in extracts of the wild type (Table 2). We consider the presence of approximately wild-type levels of RuBP carboxylase activity in the mutant extracts to suggest that the chloroplast enzymes were recovered to a similar extent in the various enzyme preparations. Mixing extracts of mutant and wild-type plants did not result in inhibition of wild-type activity. Thus, we ascribe the reduction in phosphatase activity either to reduced enzyme synthesis or to the synthesis of a defective enzyme.

A single recessive nuclear mutation seems to be responsible for both lack of viability of the mutant and enzyme deficiency. F₁ progeny from the WT × CS119 cross were phenotypically wild-type in the normal gas regime, and leaf extracts contained

~50% of the wild-type ratio of phosphatase to carboxylase activity (Table 2). The restoration of one-half of wild-type activity in the heterozygote is consistent with the expectation for a mutation within the structural gene for the phosphatase¹⁰. Of 646 F₂ progeny, 153 exhibited the mutant phenotype. The 3:1 segregation (χ^2 0.60; $P < 0.4$) indicates a single recessive mutation. Several thousand M5 mutant seeds failed to produce any phenotypically wild-type plants, indicating purity and stability of the genotype. The co-segregation of phenotype and enzyme deficiency was tested by scoring 85 randomly chosen F₂ plants for a high CO₂ requirement at an early stage of growth by the method previously described. The plants were then revived in high CO₂, grown to the floral initiation stage, and assayed for phosphatase activity. We observed an absolute correspondence between low phosphatase activity and lethality (data not shown). Thus, both effects seem to be due to the same mutation, which we have designated *pcoA-1*.

Most of the properties of the *pcoA* mutant can be explained on the basis that phosphoglycolate is a potent inhibitor of triose phosphate isomerase¹¹. When the mutant is illuminated in normal atmospheric conditions, phosphoglycolate produced by the RuBP oxygenase reaction^{12,13} may accumulate at levels which inhibit the conversion of glyceraldehyde-3-phosphate (GAP) to dihydroxyacetone phosphate. This would reduce synthesis of RuBP and concomitantly reduce photosynthesis. The restoration of photosynthetic capacity in the mutant following a brief dark treatment is attributed to the dissipation of accumulated phosphoglycolate by residual phosphatase activity. Because GAP is a major export product of photosynthesis¹⁴, inhibition of the isomerase would result in increased export of GAP as a substrate for cytoplasmic sucrose synthesis. We attribute the increased labelling of the neutral fraction in the mutant (Table 1) to such a mechanism. Also, increased labelling of the alanine pool in the mutant suggests that a proportion of the GAP may be converted to pyruvate by glycolysis.

Several properties of the mutant are directly relevant to the continuing controversy regarding the source of photorespiratory glycolate and CO₂ (refs 5, 15). First, it is now clearly established that phosphoglycolate is the major, if not sole, precursor of glycolate, as shown from an experiment in which glycolate metabolism was blocked by feeding the plants hydroxybutyric acid (HBA), a potent inhibitor of glycolate oxidase (Table 1)¹⁶. The pattern of ¹⁴CO₂ incorporation in the wild-type in the presence of HBA is consistent with similar studies with other species^{3,17} in that a large proportion of the label is found in glycolate, associated with a corresponding decrease in serine and glycine labelling. In contrast, the HBA-treated mutant does not accumulate significant amounts of labelled glycolate. Second, the low level of CO₂ evolution in the light and the absence of a post-illumination CO₂ burst in the mutant indicates that phosphoglycolate is a major precursor of photorespiratory CO₂. Third, reduced labelling of the glycine and serine pools supports the proposed flow of photorespiratory carbon through phosphoglycolate to glycolate, glycine and serine².

We thank R. Askel, D. Nash, G. Kraminsky and R. H. Hageman for technical or material assistance during this work. C.R.S. was supported by a National Research Council of Canada postdoctoral fellowship.

C. R. SOMERVILLE
W. L. OGREN

Department of Agronomy,
University of Illinois
and

US Department of Agriculture,
Science and Education Administration—Agricultural Research,
Urbana, Illinois 61801

Received 13 June; accepted 11 July 1979.

- Laing, W. A., Ogren, W. L. & Hageman, R. H. *Pl. Physiol.* **54**, 678–685 (1974).
- Tolbert, N. E. & Ryan, F. J. in *CO₂ Metabolism and Plant Productivity* (eds Burris, R. H. & Black, C. C.) 141–159 (University Park Press, Baltimore, 1976).

Table 2 Phosphoglycolate phosphatase activity in wild-type and mutant leaf extracts

Strain	Phosphoglycolate phosphatase activity*	RuBP carboxylase activity*
Wild type	0.22	0.30
CS119	0.01	0.28
F ₁ (WT × CS119)	0.08	0.23
Mixture†	0.11	0.29

Leaf tissue from 8–10 plants was ground at 4 °C in 10 mM HEPES (pH 7.0) and clarified by centrifugation at 30,000g for 20 min. Phosphatase activity was determined in 40 mM cacodylate (pH 6.3), 5 mM HEPES, 5 mM citrate, 5 mM ZnSO₄ and 0.5 mM EDTA⁹. The reaction was initiated with 2 mM phosphoglycolate dissolved in the same buffer. After 5 min at 25 °C, the reaction was terminated with acid molybdate reagent¹⁸ and the release of phosphate determined colorimetrically. RuBP carboxylase activity was determined in anaerobic conditions by measuring the incorporation of ¹⁴CO₂ into acid-stable products¹. The reaction mixture contained 50 mM Bicine (pH 8.1), 10 mM MgCl₂, 0.1 mM EDTA, 10 mM NaHCO₃ (0.1 μ Ci μ mol⁻¹), 0.4 mM RuBP and 1 mM HEPES. The reaction was initiated with enzyme activated by preincubation in a similar mixture (no RuBP), and terminated after 2 min at 25 °C. All assays were carried out in triplicate on two independent extracts. Protein was determined colorimetrically with the Folin–Ciocalteu reagent.

* Enzyme activity is expressed as μ mol per min per mg protein.

† Equal volumes of wild-type and mutant extract were mixed immediately before assay.

3. Servaites, J. C. & Ogren, W. L. *Pl. Physiol.* **59**, 587-590; **60**, 461-466 (1977).
4. Rédei, G. P. in *Genetic Manipulations with Plant Material* Vol. 3, (ed. Ledoux, L.) 329-350 (Plenum, New York, 1975).
5. Chollet, R. & Ogren, W. L. *Bot. Rev.* **41**, 137-179 (1975).
6. Chapman, E. A. & Graham, D. *Pl. Physiol.* **53**, 879-885 (1974).
7. Berry, J. A., Osmond, C. B. & Lorimer, G. H. *Pl. Physiol.* **62**, 954-967 (1978).
8. Canvin, D. T., Lloyd, N. D. H., Fock, H. & Przybylla, K. in *CO₂ Metabolism and Plant Productivity* (eds Burris, R. H. & Black, C. C.) 161-176 (University Park Press, Baltimore, 1976).
9. Christeller, J. T. & Tolbert, N. E. *J. biol. Chem.* **253**, 1780-1785 (1978).
10. MacIntyre, R. J. & O'Brien, S. J. *A. Rev. Genet.* **10**, 281-318 (1976).
11. Anderson, L. E. *Biochim. biophys. Acta* **235**, 237-244 (1971).
12. Bowers, G., Ogren, W. L. & Hageman, R. H. *Biochem. biophys. Res. Commun.* **45**, 716-722 (1971).
13. Andrews, T. J., Lorimer, G. H. & Tolbert, N. E. *Biochemistry* **12**, 11-23 (1973).
14. Fliege, R., Flügg, U.-I., Werdan, K. & Heldt, H. W. *Biochim. biophys. Acta* **502**, 232-247 (1978).
15. Zelitch, I. A. *Rev. Biochem.* **44**, 123-145 (1975).
16. Jewess, P. J., Kerr, M. W. & Whitaker, D. P. *FEBS Lett.* **53**, 292-296 (1975).
17. Kumarasinghe, K. S., Keys, A. J. & Whittingham, C. P. *J. exp. Bot.* **28**, 1163-1168 (1977).
18. Ames, B. N. *Meth. Enzym.* **8**, 115-118 (1966).

Possible function of juvenile hormone-dependent protein in larval insect diapause

JUVENILE HORMONE (JH) has been shown to control the synthesis and storage of a diapause-associated protein (DAP) in the larval fat body of the southwestern corn borer, *Diatraea grandiosella* Dyar¹. Our results indicate that DAP is synthesised *de novo* in pre-diapausing southwestern corn borers, reaches a maximum titre in newly diapaused larvae, and is used gradually during diapause. In addition, we have demonstrated recently that the protein has a molecular weight of about 35,000, and an isoelectric point of about 5.9 (unpublished). Since JH controls

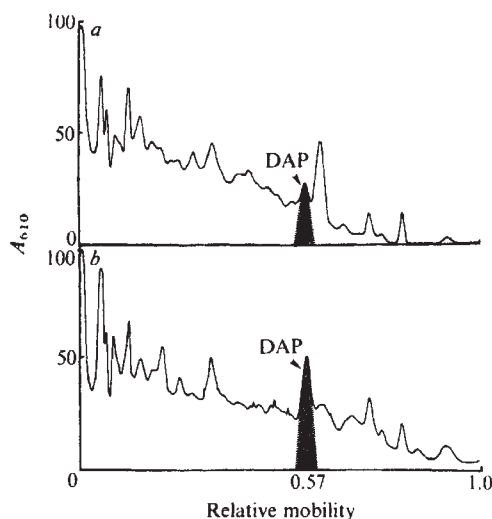


Fig. 1 Densitometric scans of electropherograms showing the effect of topical C_{17} -JH treatments on the proteins of the fat body of *D. grandiosella* larvae. *a*, Larvae (19 days old) at the beginning of the treatments. *b*, Larvae (28 days old) after 10 daily treatments ($6 \mu\text{g } C_{17}\text{-JH per } 3 \mu\text{l acetone per day}$). Larvae were reared in the diapause-averting conditions (25°C ; 16 h light, 8 h dark) until 19 days and then transferred to the diapause-inducing conditions (25°C ; 12 h light, 12 h dark). Untreated and acetone-treated controls pupated within 10 days without accumulating the diapause-associated protein (DAP), whereas those treated daily with C_{17} -JH accumulated DAP and entered a diapause-like state within about 24 days. After fat body homogenates were centrifuged at $21,000g$ for 30 min at 4°C , proteins were subjected to electrophoresis in 7% acrylamide gels using a discontinuous buffer system with Tris-glycine (pH 8.3) as the bath buffer (2 mA per gel at 4°C). Equal portions of the fat body were analysed to show changes in the total protein concentration. Based on estimates of soluble proteins present in the perivisceral fat body, the DAP content of a newly diapaused larva is about $400 \mu\text{g}$.

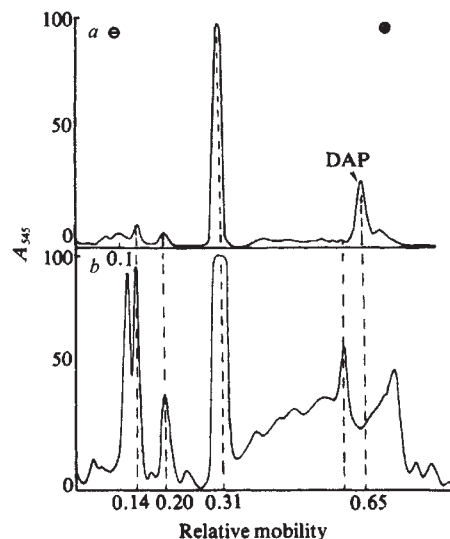


Fig. 2 Densitometric scans of electropherograms showing the relationship of proteins released from the perivisceral fat body to those present in the haemolymph in 110-day-old diapausing larvae of *D. grandiosella*. *a*, Incubation medium ($100 \mu\text{l}$, $\sim 15 \mu\text{g protein}$) after incubating one right and one left fat body sheet in $500 \mu\text{l}$ macromolecule-free Grace's medium (Grand Island Biological Co.) for 10 min. at 21°C . *b*, Haemolymph ($0.5 \mu\text{l}$, $\sim 40 \mu\text{g protein}$). Electrophoretic conditions as described in Fig. 1, except that 5% acrylamide gels were used.

the larval diapause of this insect^{2,3}, DAP may be involved in diapause maintenance. Some properties of DAP are described here and the evidence for its function is appraised. Possible functions include a nutritive store of amino acids or peptides, a reservoir of a proenzyme, or a JH-binding protein. From the available evidence, the most likely model is that of a latent store of a high affinity JH-binding protein⁴⁻⁷ which is released and activated during diapause.

The effect of C_{17} -JH on the induction of larval diapause and the accumulation of DAP was examined (Fig. 1). Larvae which were reared on an artificial diet in diapause-averting conditions long enough to programme them for metamorphosis were transferred to diapause-inducing conditions and treated with C_{17} -JH, the principal JH homologue present in diapausing larvae⁸. The treated larvae continued to feed, and about 70% moulted into pigment-free immaculate morphs and entered a diapause-like state, whereas about 90% of the untreated larvae pupated directly. Daily C_{17} -JH treatments induced the perivisceral fat bodies to accumulate substantial amounts of DAP within 10 days. Only small amounts of DAP were present in the fat body at the beginning of the JH treatments, and little, if any, DAP was present in the fat body of pupae. These findings extend earlier observations that showed a similar pattern of accumulation of DAP in the fat body of non-diapausing larvae which had been treated with a JH mimic¹. This JH regulation of protein synthesis in pre-diapausing larvae seems comparable to the JH control of vitellogenin synthesis in some adult insects⁹.

The fat body of diapausing larvae was incubated *in vitro* to study protein release (Fig. 2). Larvae were reared in diapause-inducing conditions (23°C ; photoperiod of 12 h light, 12 h dark). In these conditions they passed through six larval stages, completed feeding between 35 and 40 d, transformed into immaculate morphs, and entered diapause¹⁰. The release of DAP from the fat body of 110-day-old diapausing larvae was detected within 10 min of incubation in macromolecule-free Grace's medium. Since the release was partially inhibited by potassium cyanide (10^{-2} M) and ouabain ($5 \times 10^{-3} \text{ M}$), it appears to be an energy-requiring process. Our results also showed that DAP was more readily released from the fat body of mid-diapausing larvae (110 days old) than from that of a last stage pre-diapausing larva (35 days old), even though larvae of

both ages have a high titre of DAP in their fat bodies. In spite of the rapid discharge of DAP *in vitro*, and its gradual disappearance from the fat body of diapausing larvae *in vivo*, DAP does not accumulate in the haemolymph¹, suggesting that it may be structurally modified in the haemolymph.

To test if this structural modification of DAP occurs, incubation medium containing released DAP was treated with several concentrations of protease-containing midgut extracts. Electrophoretic examination indicated no decrease in the concentration of DAP attributable to midgut enzyme activity. Subsequently, the effect of the proteolytic enzyme, trypsin, was tested on the hydrolysis of DAP. Although incubation at 37 °C for up to 3 h resulted in the complete hydrolysis of proteins at relative mobility 0.10 and 0.20 (Fig. 2), DAP remained intact, thereby showing that DAP is relatively resistant to trypsin. Consequently, tritiated DAP was incubated for 60 min at 37 °C in a medium containing haemolymph from pre-diapausing larvae. The medium was then subjected to electrophoresis in 5% and 7% acrylamide gels and the distribution of radioactivity in

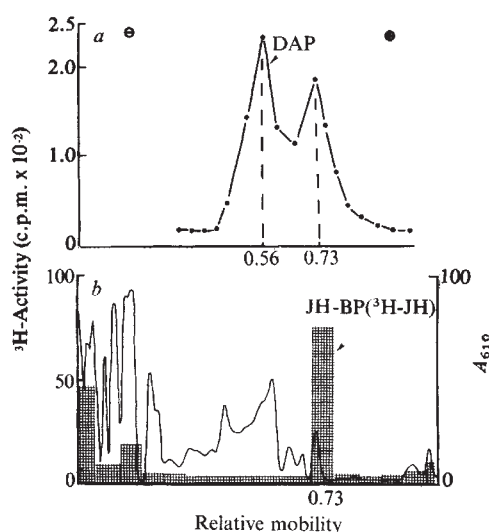


Fig. 3 Relationship between the diapause-associated protein (DAP) and the haemolymph proteins of *D. grandiosella*. *a*, Radioactivity in gel slices showing modification of a portion of prelabeled DAP following an incubation *in vitro* in macromolecule-free Grace's medium containing haemolymph from pre-diapausing larvae. Radiolabelled DAP was obtained by injecting 34-day-old pre-diapausing larvae with L-[4,5-³H(N)]leucine (3.30×10^6 d.p.m., specific activity 5.0 Ci mmol⁻¹, NEN). After 48 h fat body proteins (21,000g supernatants) were extracted and subjected to disc electrophoresis. The zone containing DAP was homogenised in distilled water and centrifuged. The supernatant was lyophilised and tritiated DAP (200 µg per 100 µl distilled water) was incubated with Grace's medium (350 µl) and haemolymph (50 µl) from 36-day-old pre-diapausing larvae for 60 min at 37 °C. The medium was electrophoresed and the gels were sliced into 2–4-mm sections. Proteins were eluted with Soluene-350 and after a stable background had been reached the radioactivity in the slices was counted in Aquasol-2 (10 ml per vial) using a Beckman LS-100C liquid scintillation counter. The formation of the modified fraction seems to depend on the concentration of haemolymph and the incubation temperature, decreasing markedly at 23 °C. *b*, Densitometric scan of electropherogram and the distribution of radioactivity in gel slices showing *in vitro* binding of tritiated C₁₇-JH to haemolymph proteins of 36-day-old pre-diapausing larvae. Haemolymph (50 µl) was preincubated for 10 min in 450 µl macromolecule-free Grace's medium containing dichlorvos (10^{-4} M) as an esterase inhibitor. The medium was then charged with [10-³H(N)]-C₁₈-JH (1 µCi, specific activity 13.5 Ci mmol⁻¹, NEN), and incubated for 20 min at 37 °C. After electrophoresis was carried out in 7% acrylamide the gels were sliced, and the radioactivity in the slices was measured as described under *a* (hatched boxes). Solid line, densitometric scan of electropherogram. JH-BP, Juvenile hormone-binding protein.

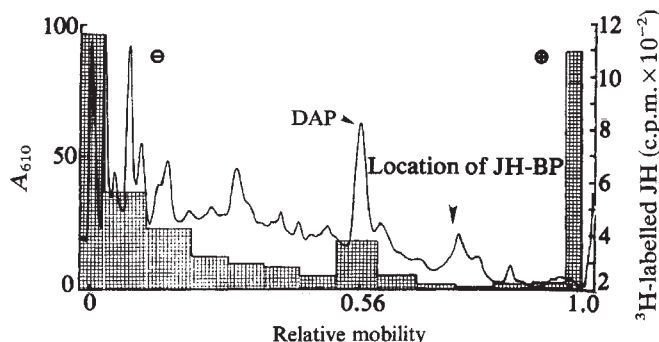


Fig. 4 Densitometric scan at 610 nm of fat body proteins (150 µg) of 39-day-old newly diapaused female larvae of *D. grandiosella* showing the binding of C₁₈-JH to protein fractions. The fat body homogenate was centrifuged at 21,000g for 30 min. The clear supernatant (300–500 µl) was treated with dichlorvos (10^{-4} M) to inhibit esterase activity, and then incubated for 20 min at 37 °C with 1 µCi of ³H-labelled C₁₈-JH. After electrophoresis in 7% acrylamide the gels were sliced into 5-mm sections and the radioactivity was measured to 1% error as described under Fig. 3*a*. Solid line, densitometric scan; hatched boxes, radioactivity. JH-BP, Juvenile hormone-binding protein.

gel slices was examined (Fig. 3*a*). In these conditions a portion of DAP was converted to a fraction with a higher relative mobility, whereas incubation without haemolymph yielded only a single peak of radioactivity corresponding to DAP. The relative mobility of the modified fraction of DAP corresponded to that of the high affinity JH-binding protein present in the haemolymph (Fig. 3*b*). Subsequently, incubation of fat body homogenates with tritiated JH indicated that DAP showed only slight affinity for JH. No binding of JH occurred in fat body proteins corresponding to the JH-binding protein present in the haemolymph (Fig. 4). Since only enzymes present in the haemolymph appear capable of modifying DAP, the binding activity of the modified fraction has not yet been measured in the absence of haemolymph JH-binding protein.

The evidence obtained on the characteristics of DAP may be summarised as follows: (1) DAP accumulates in the fat body of pre-diapausing larvae in response to their high JH titre and is gradually used during diapause¹; (2) DAP is released intact from the fat body *in vitro*, but it does not accumulate in the haemolymph of diapausing larvae *in vivo*; (3) DAP is resistant to hydrolysis by trypsin and midgut enzymes; (4) a portion of DAP is structurally modified during incubation *in vitro* in a medium containing haemolymph; (5) the amino acid composition of DAP bears little resemblance to that of the free amino acids present in the haemolymph¹; and (6) DAP shows only a weak affinity for C₁₈-JH *in vitro*.

With the exception of the sex-linked vitellogenin¹¹ relatively little is known of the functions of the exportable proteins of the insect fat body. Although DAP could serve as a store for amino acids or peptides to be used during diapause, such a function is doubtful. Early and mid-diapausing larvae rely mainly on reserve triglycerides¹², and mid-diapausing larvae already show considerable loss of DAP from their fat bodies¹. The molecular weight of DAP is also small compared with those of suspected haemolymph storage proteins¹¹. Electrophoretic evidence did not reveal any relationship between DAP and the corn borer's vitellogenin¹. Although a proenzyme function for DAP has not been ruled out, such a high requirement for a single enzyme seems unlikely (Fig. 1).

The following observations are relevant to the possible function of DAP as a latent store of JH-binding protein. Insects which undergo a JH-maintained diapause would seem to derive an adaptive advantage by storing an inactive JH-binding protein. The storage protein would not sequester JH, yet it would be readily available for release into the haemolymph during diapause when protein synthesis is at a minimum^{13,14}, thereby

ensuring the maintenance of a sufficient titre of JH in the haemolymph. The JH-binding protein present in the haemolymph of last instar pre-diapausing larvae of *D. grandiosella* has a molecular weight of 28,000, an isoelectric point of approximately 4.85, an apparent dissociation constant of approximately 3×10^{-7} M, a concentration of about 3.3×10^{-6} M, and a relative mobility of 0.73 in 7% acrylamide gels. The content of JH-binding protein in these larvae was estimated to be 11 μ g, based on a haemolymph volume of 100 μ l (unpublished). Although the half life of the binding protein is unknown, a store of about 400 μ g DAP in a newly diapaused larva (Fig. 1) would seem to fall within the limits for a JH-binding protein store required over at least a six-month period. In addition, unlike the larval fat body of two other lepidopterous insects, *Manduca sexta* and *Plodia interpunctella*^{15,16}, that of *D. grandiosella* does not seem to contain or to release an active JH-binding protein. Neither the fat body nor the midgut of the southwestern corn borer shows JH binding corresponding to the high affinity JH-binding protein, yet the available evidence implicates the insect fat body as the source of the binding protein^{15,16}. The present study suggests a critical function for a fat body protein in the maintenance of larval diapause.

This work was supported in part by NSF grant PCM78-22488 and is a contribution from the Missouri Agricultural Experiment Station. S.T. is on leave of absence from the Department of Zoology, University of Helsinki.

S. TURUNEN
G. M. CHIPPENDALE*

Department of Entomology,
University of Missouri,
Columbia, Missouri 65211

Received 14 February 1979; accepted 5 July 1979.

* To whom correspondence should be addressed.

1. Brown, J. J. & Chippendale, G. M. *Insect Biochem.* **8**, 359–367 (1978).
2. Chippendale, G. M. A. *Rev. Ent.* **22**, 121–138 (1977).
3. Yin, C.-M. & Chippendale, G. M. *J. Insect Physiol.* **25** (in the press).
4. Akamatsu, Y. *et al.* in *Control Mechanisms in Development* (eds Meints, R. H. & Davies, E.) 123–149 (Plenum, New York, 1975).
5. Kramer, K. J. & Childs, C. N. *Insect Biochem.* **7**, 397–403 (1977).
6. Goodman, W., O'Hern, P. A., Zaugg, R. H. & Gilbert, L. I. *Molec. cell. Endocr.* **11**, 225–242 (1978).
7. Klages, G. & Emmerich, H. *Insect Biochem.* **9**, 23–30 (1979).
8. Bergot, B. J., Schooley, D. A., Chippendale, G. M. & Yin, C.-M. *Life Sci.* **18**, 811–820 (1976).
9. Hagedorn, H. H. & Kunkel, J. G. A. *Rev. Ent.* **24**, 475–505 (1979).
10. Chippendale, G. M. & Yin, C.-M. *Nature* **246**, 511–513 (1973).
11. Wyatt, G. R. & Pan, M. L. A. *Rev. Biochem.* **47**, 779–817 (1978).
12. Chippendale, G. M. *Ent. exp. Appl.* **16**, 395–406 (1973).
13. Stevenson, E. & Wyatt, G. R. *Archs. Biochem. Biophys.* **99**, 65–71 (1962).
14. Hayes, D. K., Reynolds, P. S., McGuire, J. U. & Schechter, M. S. *J. Econ. Ent.* **65**, 676–679 (1971).
15. Nowock, J. & Gilbert, L. I. in *Invertebrate Tissue Culture* (eds Kurstack, E. & Maramorosh, K.) 203–212 (Academic, New York, 1976).
16. Ferkovich, S. M., Oberlander, H. & Rutter, R. R. *J. Insect Physiol.* **23**, 297–302 (1977).

Evidence that murine pre-B cells synthesise μ heavy chains but no light chains

THE first identifiable member of the B-cell line, termed the pre-B-cell, has been shown by immunofluorescent techniques to contain small amounts of intracellular IgM, but to lack detectable surface immunoglobulin¹. These cells have been found in several species at sites of B-cell generation, most notably fetal liver and adult bone marrow^{1,2} and considerable evidence now exists which supports this precursor-progeny relationship of pre-B cells and B lymphocytes^{1–5}. However, conflicting data exist with regard to pre-B cell expression of μ and light chains. Although these cells were originally described in mice as containing both intracellular heavy (μ) and light chains¹, using the same immunofluorescence techniques we have subsequently

been unable to demonstrate either κ or λ light chains within these cells (P.B. and J.K., unpublished). Moreover, the presence of light chains in murine pre-B cells cannot be confirmed with the original anti- κ antibodies used in ref. 1 (M. Cooper, personal communication). These findings are also contrary to immunochemical data which suggest the presence of complete IgM molecules both intracellularly and on the surface of early fetal liver cells⁶. We have used somatic cell hybridisation to approach this controversial issue. We have now obtained, after fusion with fetal liver, hybridomas which synthesise large amounts of intracellular μ chain in the absence of any light chain (LC) synthesis detectable by immunofluorescent or immunochemical techniques. The immunoglobulin phenotype of these hybridomas is identical to that observed in immunofluorescence studies of pre-B cells from fetal liver or adult bone marrow, and suggests that synthesis of μ chain before light chain expression is a normal event in the early differentiation of the B-cell line.

Mouse fetal liver cells were obtained from BALB/cN and C57BL/6J fetuses of different gestational ages. These cells were fused with one of three different variants of the myeloma P3 $\times$ 63Ag8 (γ 1, κ) by a modification of methods described by Lemke *et al.*⁷. These lines were (1) NS-1 ($-\kappa$) which no longer synthesises the myeloma γ 1 heavy chain but only the κ chain⁸, (2) Ag8(6), which is similar to NS-1 ($-\kappa$), and (3) Ag8653, a variant which has lost the ability to synthesise both myeloma γ 1 and κ chains even after fusion. The isolation and characterisation of the latter two cell lines are described elsewhere⁹.

The hybridomas isolated after fusion with Ag8653 are described in Table 1. It can be seen that the ratio of recoverable μ chain-synthesising hybrid cells compared with the total number of hybridomas obtained per fusion increased with the age of the fetuses, in parallel with the number of pre-B cells in fetal liver¹. Hybrids were obtained from fetal liver fusions as early as 15 days of gestation, 2 days before we can detect sIgM⁺ B cells¹⁰. The most striking observation is that of the 28 μ chain-synthesising hybridomas only 1 expressed either cytoplasmic or secreted light chains. SDS-polyacrylamide gel electrophoresis of biosynthetically labelled intracellular proteins

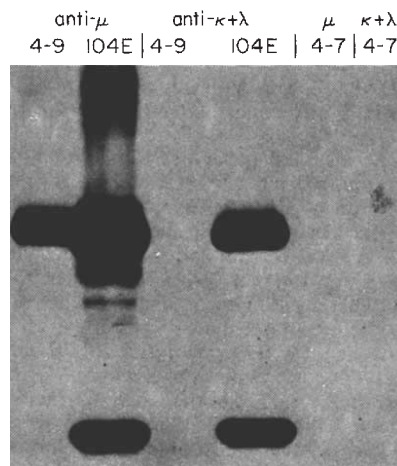


Fig. 1 Fluorogram of Ig synthesised by fetal liver hybridomas. MOPC 104E myeloma ascites cells and fetal liver hybridoma cells were cultured for 4 h in the presence of ¹⁴C- and ³H-labelled amino acids (Schwarz-Mann; protein hydrosylate), respectively. Cells were washed, lysed with 0.5% NP-40, and immunoprecipitated with purified goat antibodies to mouse μ or κ and λ chains⁷ by the *Staphylococcus aureus* technique²⁰. After elution from the bacteria with SDS-urea, the immunoprecipitates were reduced, alkylated and electrophoresed on 8–12% SDS-diallyl tartardiamide (DATD)-acrylamide gels²¹. Fluorography was by the method of Bonner and Laskey²².

Table 1 Ig-Synthesising fetal liver hybridomas derived from fusion with the non-producer myeloma Ag8653

Age* and strain of fetus	No. of fetal liver cells fused	No. of hybrids No. of wells	Immunoglobulin phenotype† of hybridomas	Frequency
15 day (C57BL/6) timed by vaginal plug	1.5×10^8	240/840	μ^+ , light chain ⁻ ; non-secretor	4/230
17 day (C57BL/6)	0.7×10^8	40/264	μ^+ , light chain ⁻ ; non-secretor	1/40
17–19 day pooled fetal livers (BALB/cN)	1.5×10^8	83/240	μ^+ , light chain ⁻ ; non-secretor	19/83
			μ^+ κ^+ ; secretor	1/83
18–19 day (BALB/cN)	0.7×10^8	30/240	μ^+ light chain ⁻ ; non-secretor	3/30

Cells were fused using polyethylene glycol (PEG) by methods described in Lemke *et al.*⁷. Briefly, dissociated fetal liver cells were mixed with an equal number of myeloma cells washed twice in pre-warmed RPMI 1640 without serum. The cell pellet was resuspended by gentle pipetting in 40% pre-warmed PEG. After ~1 min the PEG was diluted by slow dropwise addition of RPMI 1640 at 37 °C. The cells were pelleted and distributed directly in 1 ml of medium containing hypoxanthine, aminopterin and thymidine¹⁸ into 24-well Costar plates so that each well contained $\sim 5 \times 10^5$ fetal liver cells per well. Frequency = number of Ig⁺ hybridomas/total number of hybridomas.

* Unless stated the ages of the fetuses were estimated.

† Intracellular immunoglobulin was detected by two methods: direct immunofluorescence using purified antibodies to mouse μ , κ and λ chains¹⁹, or a modified enzyme-linked immunosorbent assay (ELISA). Briefly, an aliquot of hybridoma cells was detergent (NP-40) lysed *in situ* within the wells of plastic plates coated with purified goat anti-mouse μ chain antibody. Wells containing bound mouse μ chains were detected by addition of alkaline phosphatase-labelled goat anti- μ followed by the substrate *p*-nitrophenyl phosphate. A bright yellow reaction product was generated in positive wells. A more detailed description of this technique is given in ref. 9. Positive ELISA results were immediately confirmed using direct immunofluorescence. Secreted immunoglobulin was detected by analysis of tissue culture supernatants by either an ELISA, using anti- μ -coated microplates or immunoprecipitation and SDS-polyacrylamide gel analysis of biosynthetically labelled secreted protein.

immunoprecipitated from two representative fetal liver hybrids (4–9 and 4–7) is shown in Fig. 1. Two bands of radioactivity, corresponding to a μ chain and a light chain, were immunoprecipitated from the control, MOPC 104E (μ , λ plasmacytoma), by either anti- μ , or a mixture of anti- κ and anti- λ antibodies. In striking contrast, only μ chains were precipitated from 4–9 with anti- μ , and nothing with anti- κ and anti- λ . The hybridoma line 4–7 was derived from the same fetal liver fusion but was negative for immunoglobulin synthesis by immunofluorescence, a finding confirmed by biosynthetic labelling. The μ chain of 4–9 seems to be slightly larger than that of MOPC 104E on these gels. Simultaneous electrophoresis of ³H-labelled 4–9 and ¹⁴C-labelled 104E μ chains on the same gel, which was then sliced and counted after electrophoresis, confirmed this size difference (Fig. 2). The small peak of radioactivity in the light chain region of 4–9 is probably due to normal plasma cells in the peritoneal exudate feeder cells on which 4–9 was being maintained at this stage, as it has not subsequently been seen in the absence of feeder cells (data not shown). The 4–9 μ chain is also larger than the intracellular μ chain of a μ , κ hybridoma derived from adult spleen (not shown).

Although fusion of fetal liver cells with Ag8653, which contributes no heavy or light chains of its own, almost invariably resulted in hybrids with the phenotype μ^+ LC⁻ non-secretor, this did not occur when NS-1 or Ag8(6), both of which synthesize intracellular κ chains, was used. Most of the μ^+ hybridomas also synthesized κ light chains and secreted both μ and κ chains. To determine whether, in these hybrids, the light chains were derived from the myeloma or pre-B cell parent, the radiolabelled secreted IgM was isolated, reduced and separated into light and heavy chains by slab gel electrophoresis. The light chain region was cut directly from the gel and subjected to isoelectric focusing¹¹ (Fig. 3). The light chains from a fetal liver μ , κ hybridoma derived from fusion with NS-1 were identical in focusing patterns to the κ chains of the parent myeloma, P3 $\times$ 63Ag8. Therefore, the κ chains which these hybrid cells assemble with μ chains and secrete as IgM molecules are derived from the parental myeloma and no light chain is contributed by the pre-B cell.

Our results indicate that these μ^+ LC⁻ hybridomas are derived from fusion events between pre-B cells and myelomas. Alternate possibilities, that they are derived from B lymphocytes or that their phenotype is due to aberrant regulation of immunoglobulin synthesis unique to these kinds of hybridomas, are

unlikely for the following reasons. (1) The phenotype of these fetal liver hybridomas (μ^+ LC⁻) reflects that which we have observed by immunofluorescence to be normal for mouse pre-B cells. (2) We have obtained these hybridomas from fetal liver before the ontogenetic appearance of sIgM⁺ B lymphocytes. (3) There are well documented observations that the myelomas used for these fusions support synthesis of all heavy and light chain classes of immunoglobulins by hybrids made from adult spleen or lymph node cells^{9,12,13}. This mitigates against the possibility that there has been a loss of some type of regulatory mechanism from these myelomas involved in κ chain synthesis

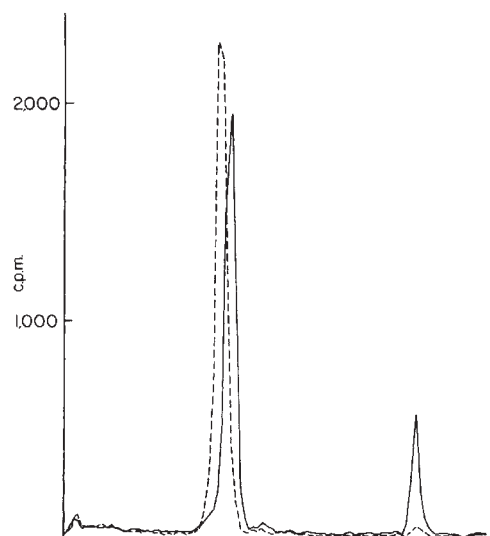


Fig. 2 SDS-polyacrylamide gel analysis of intracellular μ chains from the fetal liver hybridoma 4–9. Tritium-labelled 4–9 (---) and ¹⁴C-labelled MOPC 104E (—) cell lysates were immunoprecipitated with anti- μ as in Fig. 1 legend. Both samples were electrophoresed on the same 8–12% SDS-DATD-acrylamide gel. One-mm gel slices were dissolved in 2% periodate, and ³H and ¹⁴C counted. The 4–9 μ chain peak was 3×1 mm slices larger than the μ chain of MOPC 104E.

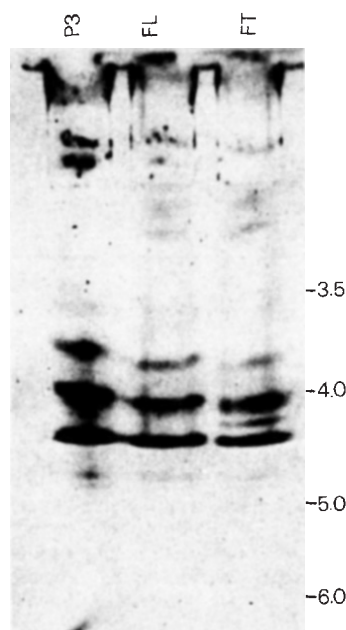


Fig. 3 Fluorogram of light chains separated by analytical isoelectric focusing in polyacrylamide gel. Purified ^{14}C - or ^3H -labelled secreted Ig was isolated as in Fig. 1, reduced and electrophoresed on 10% SDS-polyacrylamide gels. 10 μg of purified, fluorescamine-labelled, reduced and alkylated MOPC 104E IgM was included in each sample well as a marker. After electrophoresis, the light chain band was located by fluorescence of the MOPC 104E light chains on exposure of the gel to UV light. A small section (3 $\times$ 7 mm) of the gel containing the light chains was cut out and transferred to the well of an IEF gel¹¹. The gel slice was then overlaid with 6 M urea and 1% NP-40. IEF gels contained in a final concentration 4.6% acrylamide, 0.12% NN'-methylene bisacrylamide, 16.6% glycerol, 4.3 M urea, 0.5% NP-40 and 2.3% carrier ampholytes (pH 3.5–10). Polymerisation was initiated by adding 0.00016% riboflavin, 0.025% tetramethylethylenediamine and 0.00125% ammonium persulphate and exposing the casted gel to fluorescent light for 3–4 h at 4 °C. Light chains from the following cell lines are shown: (1) P3 = P3 $\times$ 63Ag8. This is the parental myeloma ($\gamma 1$, κ) from which NS-1 (ref. 8) and Ag8(6) (ref. 9) were derived. (2) FL = FL-26-13. A μ , κ fetal liver hybridoma derived from fusion of 18–19-day fetal liver cells with NS-1. Comparison with lane 1 reveals that the light chains were derived from NS-1. (3) FT = FT-93-4. A hybridoma secreting μ in association with both κ and λ light chains. The κ chains seem to be provided by NS-1, whereas the extra band probably represents the λ chains of the antibody forming cell. Tissue culture supernatants from (1) were immunoprecipitated with anti- $\gamma 1$. Those from (2) and (3) were precipitated with anti- μ .

which cannot be reconstituted by the pre-B cell alone. (4) In an examination by double immunofluorescence of over 2,000 individual cells from 24 hybridoma-containing wells derived from a fusion of Ag8653 with spleen cells from hapten-immunised C57BL/6 mice, none was found that expressed μ chains without κ or λ chains (J.K. and P.B., unpublished observations). These results also agree with other findings that normally after fusions with mouse myelomas, heavy chain loss precedes light chain loss (refs 12, 13 and J.K. unpublished). (5) In these experiments individual cells of fetal liver-derived hybridomas were screened by immunofluorescence 14–15 days after fusion, directly from the original tissue culture wells. Thus, there was little opportunity for the development of light chain negative variants. If selective chromosomal loss were the explanation, it would have had to occur simultaneously in all cells in the clone (which we have never observed) or at the time of fusion. (6) In addition, it has recently been shown that some

Abelson virus and chemically transformed cell lines believed to be derived from early B cells also express only μ heavy chains with no concomitant light chain synthesis^{14,15}. (7) Direct immunochemical analysis of radiolabelled murine fetal liver cells also supports our hypothesis that μ chain synthesis normally precedes light chain expression (D. Levitt, unpublished observations).

As we have obtained predominantly $\mu^+ \text{LC}^-$ hybridomas from fetal livers even at a time of development, when there are at least as many $\text{sIgM}^+ \text{LC}^+$ B lymphocytes present as there are pre-B cells, how are we to explain the bias towards hybridomas with a pre-B cell phenotype? We have previously shown that in fusions of lipopolysaccharide (LPS) activated B cells the maximum fusion frequency correlated with peak cell proliferation¹⁶. The finding that many pre-B cells have been shown to be rapidly dividing while most newly generated B cells are not^{3,4,17} may explain why two cell types representing such extremes of differentiation, that is, a pre-B cell and a myeloma, can fuse successfully. The immunoglobulin-negative fetal liver hybridomas (for example 4–7), which are frequently obtained, would by the same reasoning be derived from other proliferating fetal liver cells. Whether these cells are lymphoid or are derived from non-lymphoid embryonic cells, their further analysis should prove interesting.

The biological significance of asynchronous onset of μ and light chain synthesis is, of course, open to speculation but could be related to mechanisms involved in the generation of antibody diversity. If variable-region genes are expressed in an orderly, preprogrammed manner, it would be advantageous for a pre-B cell with extensive self-renewal capacity to fix a particular $\text{C}\mu \text{V}_\text{H}$ gene combination. In this sense a single pre-B cell would become the stem cell for a given V_H , and its offspring could then select from the entire V_L repertoire by a similar mechanism. In this way extensive combinatorial associations between V_H and V_L would be possible in the immunocompetent B cells generated from their pre-B cell precursors.

Because pre-B cells synthesise μ chains but apparently secrete little, if any, immunoglobulin, fusion of these cells with appropriate myelomas has allowed us to expand the clonal product of these cells. We will then be able to study pre-B cell μ chains before any kind of positive or negative selection can be exerted through the specificities of these immunoglobulins, as these cells do not seem to express stable receptors^{1,5}. For these reasons the use of hybridomas made from fetal liver pre-B cells provides a unique means by which the functions, immunoglobulin structure and specificity of pre-B cells can be studied.

We thank K. Rajewsky, A. R. Lawton, M. Cooper and D. Levitt for helpful criticism; Ms Gertrud von Hesberg and Ms Amy Anderson for technical assistance and Ms Ann Brookshire for preparation of this manuscript. Part of this work was performed while J.K. was the recipient of a Visiting Senior EMBO fellowship in the Department of Genetics, University of Cologne. This work was also supported by NIH grants CA 16673, CA 13148, AI 14782 and T32CA09128. J.F.K. is recipient of a Research Career Development Award, 1 K04 AI CA 00338.

PETER BURROWS
MARK LEJEUNE
JOHN F. KEARNEY

*The Cellular Immunobiology Unit of the Tumor Institute,
Departments of Microbiology
and
The Comprehensive Cancer Center,
University of Alabama in Birmingham,
Birmingham, Alabama 35294*

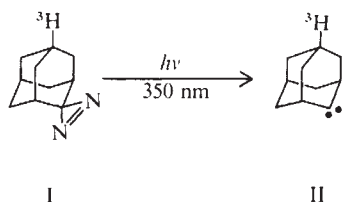
Received 12 June; accepted 20 July 1979.

1. Raff, M. D., Megson, M. Owen, J. J. T. & Cooper, M. D. *Nature* **259**, 224 (1976).
2. Gathings, W. E., Lawton, A. R. & Cooper, M. D. *Eur. J. Immun.* **7**, 804 (1977).
3. Owen, J. J. T., Wright, D. E., Habu, S., Raff, M. C. & Cooper, M. D. *J. Immun.* **118**, 2067 (1978).
4. Okos, A. J. & Gathings, W. E. *Fedn Proc.* **37**, 1770 (1977).

5. Burrows, P. D., Kearney, J. F., Lawton, A. R. & Cooper, M. D. *J. Immun.* **120**, 1526 (1978).
6. Melchers, F., Andersson, J. & Phillips, R. A. *Cold Spring Harb. Symp. quant. Biol.* **44**, 147 (1976).
7. Lemke, H., Hämmerling, G. J., Hohmann, C. & Rajewsky, K. *Nature* **271**, 249 (1971).
8. Köhler, G. & Milstein, C. *Nature* **256**, 495 (1976).
9. Kearney, J. F., Radbruch, A., Liesegang, B. & Rajewsky, K. *J. Immun.* (in the press).
10. Owen, J. J. T., Cooper, M. D. & Raff, M. C. *Nature* **249**, 361 (1974).
11. Perlmutter, R. M., Briles, D. E. & Davie, J. M. *J. Immun.* **118**, 2161 (1977).
12. Secher, D. S., Milstein, C. & Adetugbo, K. *Immun. Rev.* **36**, 51 (1977).
13. Margulies, D. H. *et al. Cold Spring Harb. Symp. quant. Biol.* **41**, 781 (1976).
14. Siden, E. J., Baltimore, Clark, D. & Rosenberg, N. E. *Cell* **16**, 389 (1979).
15. Paige, C. J., Kincade, P. W. & Ralph, P. J. *Immun.* **121**, 641 (1978).
16. Paslay, J. W., Kearney, J. F., Anderson, A. E., Bennett, J. C. & Roozen, K. J. *Fedn Proc.* **37**, 1770 (1978).
17. Osmond, D. G. & Nossal, G. V. *Cell. Immun.* **13**, 132 (1974).
18. Littlefield, J. W. *Science* **145**, 709 (1964).
19. Kearney, J. F. & Lawton, A. R. *J. Immun.* **115**, 671 (1975).
20. Kessler, S. J. *Immun.* **115**, 1617 (1975).
21. Laemmli, U. K. *Nature* **222**, 650 (1970).
22. Bonner, W. M. & Laskey, R. A. *Eur. J. Biochem.* **46**, 83 (1974).

Selective labelling of the hydrophobic segments of intrinsic membrane proteins with a lipophilic photogenerated carbene

HYDROPHOBIC photogenerated reagents are attractive candidates for the labelling of those parts of membrane proteins that are buried within the lipid bilayer. Recently, nitrenes generated from aryl azides have been used to distinguish intrinsic from extrinsic membrane proteins¹⁻⁴ and to label glycoporphin A in or close to the transmembrane hydrophobic sequence⁵. We have questioned the use of aryl nitrenes on the basis of their evident lack of reactivity towards carbon-hydrogen bonds in



single-bilayer phospholipid vesicles⁶. Because the lipid binding sites of intrinsic membrane proteins are believed to contain chemically unreactive hydrophobic amino acids, a reactive species that is less selective than a nitrene is needed to increase the probability that all the protein segments that penetrate the lipid bilayer will be labelled. In general, carbenes are more reactive than nitrenes (ref. 7 and refs therein); indeed, we have shown that carbenes generated within lipid bilayers do insert into saturated fatty acyl chains⁸. It is important to characterise the reactivity of lipophilic photogenerated reagents with well studied membrane proteins, whose membrane-associated regions have been identified by other experiments. We report here that the carbene, ³H-adamantylidene (II), generated photochemically within biological membranes from ³H-adamantane diazirine (I), labels three such intrinsic membrane proteins in or close to the hydrophobic peptides that are believed to lie within the hydrocarbon region of the lipid bilayer.

The carbene precursor 1-[³H]spiro[adamantane-4,4'-diazirine] (I) is a stable compound which absorbs in the near UV ($\lambda_{\max}$ 372.5 nm, ϵ 245; the previously reported value⁸ of 167 is incorrect). In the presence of phospholipid vesicles, the concentration of the diazirine within the bilayer is about 6,000-fold higher than the concentration in the surrounding solution⁹. Photogenerated ³H-adamantylidene, which unlike most other dialkyl carbenes does not rearrange to unreactive molecules, labels the intrinsic proteins of the human erythrocyte membrane far more heavily than the extrinsic proteins⁹. Details of the

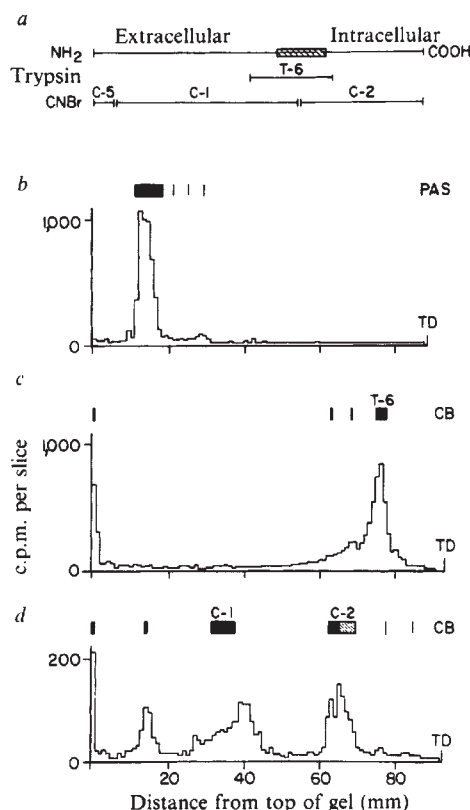


Fig. 1 Location of the sites in glycoporphin A labelled by ³H-adamantylidene. *a*, Schematic diagram of glycoporphin A indicating the hydrophobic region (shaded), the T-6 tryptic peptide and the cyanogen bromide fragments (C-5, C-1 and C-2)^{12,13}. *b*, SDS-PAGE profile of purified, labelled glycoporphin A. ³H-adamantane diazirine in ethanol was slowly added to a stirred suspension of human erythrocyte membranes (3–4 mg ml⁻¹ of protein) in a glass-walled tube. The membrane preparation had been thoroughly purged with N₂ before the addition. The final concentration of the reagent was 12 μ M (10 μ Ci ml⁻¹, specific radioactivity 840 mCi mmol⁻¹) and of ethanol, 1.0–1.5% (v/v). The solution was incubated on ice for at least 30 min, and then it was photolysed while stirring for 15 min at $\sim$ 10 $^{\circ}$ C, 3 cm from a Rayonet RPR 3,500 $\text{\AA}$ UV lamp. The labelled membranes were diluted 10-fold into bovine serum albumin (10 mg ml⁻¹) in 5 mM Na₂HPO₄, pH 8.0, allowed to stand for 10–15 min at 4 $^{\circ}$ C and then recovered by centrifugation. After four such washes, the membranes were washed four times with 5 mM Na₂HPO₄, pH 8.0. The final pellet contained 35–40% of the radioactivity added initially. Glycoporphin A was purified from the washed membranes by the LIS-phenol method¹⁴, followed by gel filtration on Sepharose 6B-CL (Pharmacia) in 0.1% Ammonyx-LO (Onyx Chemical Co.), 5 mM Na₂HPO₄ and 25 mM NaCl, pH 7.4 (ref. 15). *c*, SDS-PAGE profile of the hydrophobic peptide, T-6. Purified glycoporphin A was treated with trypsin and the T-6 peptide was recovered by acid precipitation. The radioactivity co-migrates with the broad Coomassie Blue-stained band characteristic of the T-6 fragment^{11,16}. *d*, Distribution of radioactivity and Coomassie Blue stain after SDS-PAGE of the fragments generated by treating glycoporphin A with cyanogen bromide for 48 h (ref. 12). All samples were subjected to electrophoresis on 12% acrylamide, 0.4% bisacrylamide tube gels. The electrophoresis buffer was 0.1 M Tris-Bicine, 0.1% SDS, pH 8.3. The gels were stained with periodic acid-Schiff's reagent (PAS) or with Coomassie Blue (CB). Stained gels were cut into 1-mm slices which were incubated in Protosol/Liquifluor/toluene (5:4:91, v/v/v) at 40 $^{\circ}$ C for 12 h and then counted in a liquid scintillation counter. The position of the tracking dye (bromphenol blue) is indicated by TD.

synthesis of the reagent and its full characterisation will be reported elsewhere.

The first intrinsic membrane protein that we have investigated in detail is glycoporphin A, the major sialoglycoprotein of the human erythrocyte, which contains a single polypeptide chain that spans the plasma membrane¹⁰⁻¹¹. It has been proposed that residues 73–95, which form a predominantly hydrophobic sequence of amino acids, lie within the bilayer¹⁰. Glycoporphin A was purified from human erythrocyte ghosts that had been labelled with ³H-adamantylidene. The distribution of periodic acid-Schiff's reagent stain, Coomassie Blue stain (not shown)

and radioactivity on SDS-polyacrylamide gel electrophoresis (PAGE) showed that the isolated glycoprotein was more than 95% pure (Fig. 1b). The hydrophobic peptide (T-6, Fig. 1a) that spans the bilayer was isolated from the labelled glycoprotein A by trypsin treatment followed by acid precipitation^{11,16}. Most of the radioactivity (70–85%) was recovered in the precipitate which gave a single broad band of radioactivity on SDS-PAGE that corresponded to the Coomassie Blue-stained T-6 peptide^{11,12} (Fig. 1c). The small amount of radioactivity that did not precipitate with the T-6 peptide could not be traced to a specific region of the glycoprotein molecule. In particular, the sialoglycopeptides contained no detectable radioactivity when analysed by electrophoresis. The distribution of label within glycoprotein A was unchanged when labelling was carried out in the presence of glutathione (40 mM), a water-soluble scavenger of reactive intermediates^{6,8}.

The cyanogen bromide fragments of glycoprotein A, C-1 and C-2, that overlap the hydrophobic segment^{12,13} were approximately equally labelled (Fig. 1d). Adamantylidene therefore reacts at more than one site in glycoprotein A. Furthermore, the overlap between the C-terminal fragment, C-2, and the hydrophobic peptide, T-6, represents (except for Arg 96 and Arg 97) 14 of the hydrophobic residues¹³, suggesting that the label in C-2 arises from attack by adamantylidene from within the lipid bilayer. The overlap between C-1 and T-6 contains the 11 hydrophilic residues at the N-terminus of T-6 in addition to 9 hydrophobic residues¹³. It is therefore possible that part of the label in C-1 represents labelling of hydrophilic residues just outside the hydrophobic sequence.

In a second study, we have labelled the human major histocompatibility antigens (HLA-A2 and HLA-B7), which are complexes of glycosylated transmembrane heavy chains [molecular weight (MW) 44,000] and extracellular β_2 -microglobulin molecules (MW 12,000)¹⁷. The amino acid sequences of the heavy chains of HLA-A2 and HLA-B7 are highly homologous (H.T. Orr, personal communication, and ref. 18): each molecule has a large N-terminal, extracellular sequence (~280 residues), a penultimate hydrophobic sequence (26 residues) and a C-terminal hydrophilic sequence (32 residues)^{19,20}. As the C-terminus is intracellular²¹, the hydrophobic sequence is likely to lie in the interior of the membrane. Papain cleaves native, detergent-solubilised HLA-A2 and HLA-B7 antigens in two steps²². In the first step, most of the C-terminal hydrophilic sequence is removed. In the second step, the hydrophobic sequence and a small number of the adjacent amino acids are cleaved from the large hydrophilic N-terminal domain (Fig. 2a). When HLA-A2 and HLA-B7 in lymphoblastoid cell membranes were labelled with ³H-adamantylidene, tritium was incorporated into the heavy but not the light chains. Labelled, purified HLA-A2 and HLA-B7 were treated with papain (Fig. 2b–d). The first cleavage product carried ~85% and the second cleavage product <15% of the label found in the uncleaved heavy chain. The label must therefore lie in or close to the transmembrane sequence of the HLA-A2 and HLA-B7 antigens.

Finally, we have examined the labelling of influenza virus haemagglutinin, HA₂, the major intrinsic protein of the viral envelope^{26,27}. When influenza virus is treated with bromelain, HA₂ (MW 25,000) is cleaved, yielding a water-soluble fragment, BHA₂ (MW 20,000), and a membrane-bound 'tail' (refs 28, 29). Intact influenza virus (JAP/305/1957) was labelled with ³H-adamantylidene. About one-fifth of the protein-bound label was found in the matrix protein (M) and the remainder in the haemagglutinin subunit, HA₂. Little or no label was found in the neuraminidase, the nucleocapsid protein, or HA₁, which is attached to the outside of the virion by disulphide linkages to HA₂ (ref. 30). When the labelled, washed virus was treated with bromelain, at least 75% of the radioactivity in HA₂ was not present in BHA₂. Electrophoresis of the bromelain-digested, rewashed virus showed that a small radioactive peptide (MW ~5,000) was retained by the membranes³¹. Similar results have been obtained with a nitrene precursor²⁹.

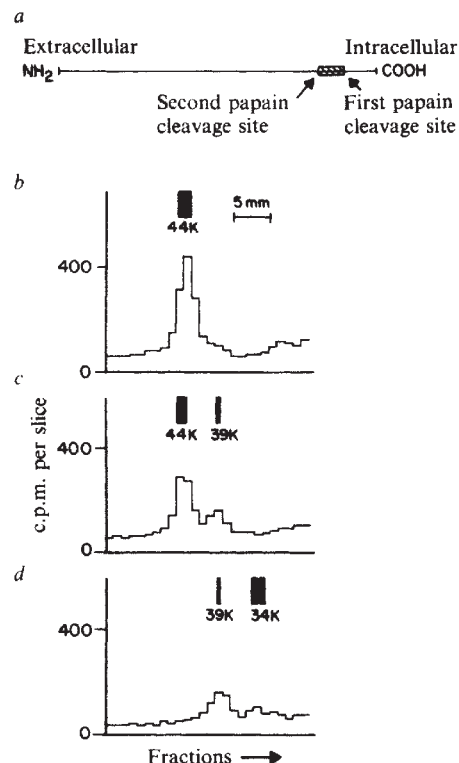


Fig. 2 Location of the sites in the heavy chains of HLA-A2 and HLA-B7 antigens labelled by ³H-adamantylidene. *a*, Schematic diagram of the heavy chain of HLA-A2 and HLA-B7 antigens showing the approximate positions of papain cleavage and (shaded) the hydrophobic region of the polypeptide^{19,20,22}. *b–d*, Labelled HLA-A2 and HLA-B7 heavy chains and their major cleavage products analysed by SDS-PAGE. Membranes, prepared from JY lymphoblastoid cells (1 g, homozygous HLA-A2 and HLA-B7)^{23,24}, were suspended in phosphate-buffered saline (0.1 ml), and labelled as described in Fig. 1 (but in air) with 1.0 mCi of ³H-adamantane diazine (final ethanol concentration 1.5%). The labelled membranes were washed twice with phosphate-buffered saline and extracted into 2% Nonidet P40 (ref. 21). The extract was adsorbed onto rabbit anti-human β_2 -microglobulin IgG coupled to Sepharose beads, as described previously²¹. Immunoadsorption both purifies and concentrates the antigens. Aliquots of HLA-A2 and HLA-B7 (10 μ g in 150 μ l of buffer) eluted from the beads with β_2 -microglobulin, were subjected to proteolysis with various amounts of papain for 1 h at 37 °C. The products were analysed by SDS-PAGE²⁵. 80% of each sample was subjected to electrophoresis on tube gels which were subsequently fractionated into 1-mm slices with a Gilson Aliquogel apparatus. The slices were minced in 2% SDS, incubated with 10 ml of Biosolve/Liquifluor/Toluene (10:5:85 v/v/v) and assayed by liquid scintillation counting. The remainder of each sample was run on a slab gel which was stained with Coomassie Blue. Only the relevant portions of the gels are shown. *b*, (No papain) uncleaved heavy chain (MW 44,000; 100%); *c*, (0.05 μ g papain) uncleaved heavy chain (~75%), and the first cleavage product (MW 39,000; ~25%); *d*, (0.5 μ g papain) no uncleaved material, the first cleavage product (~25%) and the final cleavage product (MW 34,000, ~75%). The doublet of the final product is caused by separation of the HLA-A2 and HLA-B7 specificities. Fluorography of the slab gels confirmed the exact coincidence of the radioactive peaks with the Coomassie Blue-stained bands. The apparent MWs of the antigens and their fragments are probably not accurate as their differences do not correspond to the sizes of the fragments released on proteolysis^{19,20,22}.

It is clear from the present results that adamantylidene is an effective reagent for the identification of hydrophobic, membrane-bound segments of intrinsic membrane proteins. The greater reactivity of the carbene offers a significant advantage over the more commonly used lipophilic nitrenes. However, several points remain to be determined. First, the label within the hydrophobic sequences may be attached only to the more reactive amino acids (for example, Thr, Ser and Tyr in the case of glycoprotein A). Second, we have not shown whether all the label is located within the hydrophobic sequences of the labelled proteins. Part of the label may lie just outside those

sequences. Even carbenes exhibit some chemical selectivity (ref. 7 and refs therein) and the rate of reaction of adamantylidene with the rather inert groups within the bilayer might be slower than diffusion to the membrane surface and reaction with the nucleophilic groups present near the interface. Furthermore, we have shown previously that, when irradiated, adamantane diazirine rearranges in part to 2-diazoadamantane⁸. It is possible that some of the labelling that we have observed is derived from this reactive molecule. It is interesting that the neuraminidase of influenza virus, which is regarded as an intrinsic protein³², is barely labelled compared with HA₂ even when the higher number of copies per virion of the latter³³ is taken into account. It may be that the hydrophobic region of the neuraminidase is unreactive towards the carbene, that the hydrophobic region is shielded by tightly bound lipid or protein, or that this protein is actually only indirectly attached to the bilayer.

The three intrinsic proteins that we have studied are of a common structural type. They each contain a single hydrophobic peptide that interacts with the lipid bilayer. When the sites of labelling by adamantylidene have been elucidated at the level of individual amino acids, we will be able to use the reagent for structural studies of more complex intrinsic proteins (for example, those with polypeptide chains that span the membrane more than once) and membrane proteins of completely unknown structure. Even at the present stage of development, labelling with adamantane diazirine is useful as a convenient and rapid method for the preliminary characterisation of membrane proteins.

We are grateful for the advice of Professors J. R. Knowles, G. Guidotti, J. L. Strominger and D. Wiley in whose laboratories this research was carried out. The work was supported by grants from the NIH and the NSF. J.S.P. is the recipient of a fellowship from the Anna Fuller Fund.

DANIEL W. GOLDMAN
JORDAN S. POBER

*The Biological Laboratories,
Harvard University*

JUDY WHITE

*Gibbs Laboratory,
Harvard University*

HAGAN BAYLEY*

*Mallinckrodt Laboratory,
Harvard University,
12 Oxford Street,
Cambridge, Massachusetts 02138*

Received 28 February; accepted 2 July 1979.

* To whom correspondence should be addressed at Room 18-510, Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

- Klip, A. & Gitler, C. *Biochem. biophys. Res. Commun.* **60**, 1155-1162 (1974).
- Bercovici, T., Gitler, C. & Bromberg, A. *Biochemistry* **17**, 1484-1489 (1978).
- Karlsh, S., Jørgensen, P. L. & Gitler, C. *Nature* **269**, 715-717 (1977).
- Sigrist-Nelson, K., Sigrist, H., Bercovici, T. & Gitler, C. *Biochim. biophys. Acta* **468**, 163-176 (1977).
- Kahane, I. & Gitler, C. *Science* **201**, 351-352 (1978).
- Bayley, H. & Knowles, J. R. *Biochemistry* **17**, 2414-2419 (1978).
- Bayley, H. & Knowles, J. R. *Meth. Enzym.* **46**, 69-144 (1977).
- Bayley, H. & Knowles, J. R. *Biochemistry* **17**, 2420-2423 (1978).
- Bayley, H. thesis Harvard Univ. (1979).
- Segrest, J. P., Jackson, R. L., Marchesi, V. T., Guyer, R. B. & Terry, W. *Biochem. biophys. Res. Commun.* **49**, 964-969 (1972).
- Furthmayr, H., Galardy, R. E., Tomita, M. & Marchesi, V. T. *Archs Biochem. Biophys.* **185**, 21-29 (1978).
- Segrest, J. P., Kahane, I., Jackson, R. L. & Marchesi, V. T. *Archs Biochem. Biophys.* **155**, 167-183 (1973).
- Tomita, M., Furthmayr, H. & Marchesi, V. T. *Biochemistry* **17**, 4756-4770 (1978).
- Marchesi, V. T. & Andrews, E. P. *Science* **174**, 1247-1248 (1971).
- Furthmayr, H., Tomita, M. & Marchesi, V. T. *Biochem. biophys. Res. Commun.* **65**, 113-121 (1975).
- Jackson, R. L., Segrest, J. P., Kahane, I. & Marchesi, V. T. *Biochemistry* **12**, 3131-3138 (1973).
- Springer, T. A., Strominger, J. L. & Mann, D. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1539-1543 (1974).
- Terhorst, C., Parham, P., Mann, D. L. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **73**, 910-914 (1976).
- Robb, R. J., Terhorst, C. & Strominger, J. L. *J. biol. Chem.* **253**, 5319-5324 (1978).
- Robb, R. J. thesis, Harvard Univ. (1978).
- Pober, J. S., Guild, B. C. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **75**, 6002-6006 (1978).
- Springer, T. A. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2481-2485 (1976).
- Turner, M. J. *et al. J. biol. Chem.* **250**, 4512-4519 (1975).
- Robb, R. J., Strominger, J. L. & Mann, D. L. *J. biol. Chem.* **251**, 5427-5428 (1976).
- Laemmli, U. K. *Nature* **227**, 680-685 (1970).
- Schulze, I. T. *Adv. Virus Res.* **18**, 1-55 (1973).
- Laver, W. G. *Adv. Virus Res.* **18**, 57-103 (1973).
- Brand, C. M. & Skehel, J. J. *Nature new Biol.* **238**, 145-147 (1972).
- Wiley, D. & Skehel, J. in *Topics in Infectious Diseases* Vol. 3 (eds Laver, W. G., Bachmayer, H. & Weil, R.) 135-138 (Springer, Vienna, 1978).
- Skehel, J. & Waterfield, M. *Proc. natn. Acad. Sci. U.S.A.* **72**, 93-97 (1975).
- White, J. thesis Harvard Univ. (1979).
- Laver, W. G. & Valentine, R. C. *Virology* **38**, 105-119 (1969).
- Bucher, D. & Palese, P. in *The Influenza Virus and Influenza* (ed. Kilbourne, E. D.) 83-123 (Academic, New York, 1975).

A novel coenzyme from bacterial primary alcohol dehydrogenases

METHYLOTROPHIC bacteria are able to use methane derivatives as their sole source of carbon and metabolic energy and so can sustain growth on methane, methanol and other organic compounds which lack carbon-carbon bonds¹. They are not autotrophic and, being unable to use carbon dioxide, rely ultimately on their ability to oxidise these substances. Although the primary alcohol dehydrogenases from a variety of methanol-grown bacteria have been purified and compared, and despite the increasing commercial importance of these organisms as sources of 'single cell protein', the essential features of these oxidative pathways remain unknown. The dehydrogenases exhibit a rather broad substrate specificity for primary alcohols and formaldehyde, and seem to contain a common and novel cofactor²⁻⁶, which may be more generally associated with the oxidation of single-carbon compounds. The structure of the cofactor is reported here.

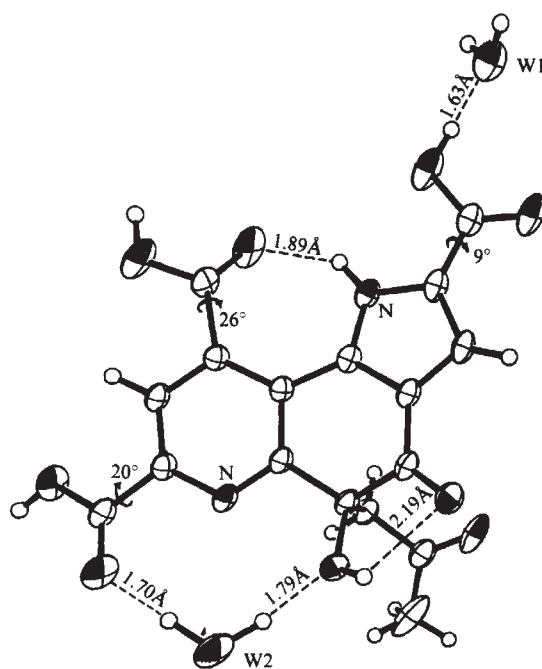


Fig. 1 A view of the methoxatin molecule¹⁰ in the same orientation as the chemical diagram. Two water molecules are also shown. O atoms are shaded, N atoms shaded and labelled. Dotted lines indicate hydrogen bonds, and arrows the sense of the torsion angles marked on the diagram.

The presumptive cofactor, obtained by denaturing the holoenzymes by heat or addition of acid, base, potassium bromide or organic solvents⁵, slowly but spontaneously decomposes in water⁷, a process which is apparently accelerated by the addition of acetone. To characterise the cofactor, which was presumed to be a pterin⁸, we devised a procedure for isolating this degradation product from intact bacteria using, as the best source, cell paste of the facultative methylotroph *Pseudomonas* TP1 (ref. 9) grown on methanol. On treating aqueous solutions of the partially purified cell extract with equal volumes of acetone a product was obtained which, although very sensitive to base, could be separated by ion exchange chromatography using acidic solutions, in which it was almost indefinitely stable. The pale yellow solid isolated in this way in a yield of 5 mg per kg cell paste was identical to the material obtained from homogeneous dehydrogenase preparations after similar acetone treatment. It crystallised as effluorescent orange needles from supersaturated aqueous solution.

The limited quantities of material available and the difficulties encountered in achieving specific degradations made this an unsuitable approach to the elucidation of the compound's chemical structure, and spectroscopic data alone provided insufficient information. The structure could, however, be determined by X-ray diffraction analysis using two single crystals.

The crystals were triclinic, space group $P\bar{1}$ with unit cell dimensions: $a = 8.593(1)$, $b = 10.702(2)$, $c = 10.265(2)$ Å; $\alpha = 101.78(2)$, $\beta = 107.77(2)$, $\gamma = 91.48(2)^\circ$. By assuming a molecular weight of 366, suggested by rather equivocal mass spectra, and $Z = 2$, the density calculated for a dihydrate, $D_c = 1.61$ g cm⁻³, compared well with that obtained by flotation measurements on the crystal, $D_m = 1.61$ g cm⁻³. Using a crystal mounted in a thin-walled glass capillary, diffractometer data were collected for 2,982 independent reflections with $2\theta_{\max} = 130^\circ$ and CuK α radiation. The intensities were empirically corrected for absorption and processed in the conventional manner.

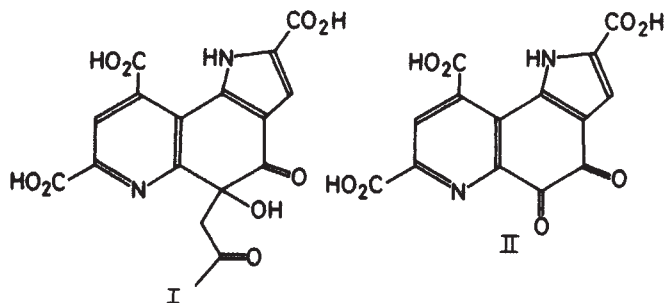
The structure was solved by direct methods using a locally modified version of SHELX76 (G. M. Sheldrick, personal communication). A set of 28 peaks in the original E map which were within chemically reasonable bonding distance were assigned as carbon atoms and later identified (as C, O or N) by the interpretation of Fourier difference maps and refinement of the thermal parameters. All hydrogen atoms were unambiguously located from difference maps during the initial cycles of anisotropic refinement. The final reliability factor, with hydrogen atoms refined isotropically and other atoms anisotropically, was $R = 4\%$.

The structure obtained from this analysis is shown in Fig. 1. The cofactor derivative is thus 4,5-dihydro-5-hydroxy-4-oxo-5-(2-oxopropyl)-1H-pyrrolo[2,3-f]quinoline-2,7,9-tricarboxylic acid (I) and crystallises as the dihydrate with the formula $C_{17}H_{12}N_2O_9 \cdot 2H_2O$ (formula weight 424).

This structure has an unusual network of hydrogen bonds. There is a strong intramolecular hydrogen bond between the pyrrolo-nitrogen atom and the carbonyl oxygen atom of the 9-carboxy substituent. A weaker bifurcated intramolecular hydrogen bond occurs between the 5-hydroxy and the 4-keto substituents. The two water molecules have very different roles in the structure: whereas W1 is important in maintaining the crystal integrity by participating in three intermolecular hydrogen bonds with symmetry-related molecules, W2 forms a tight intramolecular hydrogen-bonded ring by bridging the keto oxygen atom of the 7-carboxy-substituent and the hydroxyl oxygen atom.

Lack of structural information about this cofactor has probably contributed to the slow progress of studies of bacterial primary alcohol dehydrogenases. The crucial relationship between the structure of the artifact I, presented here, and that of the coenzyme itself must remain equivocal until the identity of the latter can be established by reconstitution of the holoenzyme from apoenzyme and cofactor. However, the comparatively

mild and diverse conditions which bring about the reverse of this process and concomitant inactivation suggest that the green-fluorescing compound obtained by denaturing aqueous solutions of the dehydrogenases differs, if at all, only very slightly from the actual cofactor. The formation of I from this compound may be more straightforward; the presence of a 2-ketopropyl moiety at a racemic centre (a consequence of the space group) points strongly to a non-biological origin. We believe it arises during the isolation from the addition of acetone to the orthoquinone (II), 4,5-dihydro-4,5-dioxo-1H-pyrrolo[2,3-f]quinoline-2,7,9-tricarboxylic acid, for which we propose the trivial name 'methoxatin'. The similar appearance of the product without acetone might then be explained by hydration which would alter the chromophore in an almost identical manner.



Methoxatin, although it is a quinone, cannot be simply categorised as such because other parts of this novel and surprisingly complex molecule are likely to contribute directly to its presumed function as a catalyst in biological oxidations. Apart from the relative orientations of the nitrogen atoms and carbonyl groups, the presence of three carboxylic acid functions is also remarkable and offers a ready explanation for the difficulties encountered in reconstituting the holoenzyme.

Experiments to reconstitute the holo-dehydrogenase and so test the view that II is indeed the cofactor, together with a total synthesis of this compound, are now in progress. Details of the isolation and properties of I, and a discussion of the possible mechanism of enzymatic catalysis involving II will be presented elsewhere.

We thank Shell Research Division and ICI for supplying, respectively, spray dried cells of *Pseudomonas extorquens* and Prutein, the 'single cell protein' product from *Methylophilus methylotrophus*. We acknowledge the contribution of Dr G. Sperl to early work on isolation of the cofactor. This work was supported by NSF grant PCM 76-04988, the Robert A. Welch Foundation and the MRC.

S. A. SALISBURY*
H. S. FORREST

Department of Zoology,
University of Texas at Austin,
Austin, Texas 78712

W. B. T. CRUSE
O. KENNARD

University Chemical Laboratory,
Lensfield Road, Cambridge, UK

Received 14 June; accepted 11 July 1979.

* Present address: University Chemical Laboratory, Lensfield Road, Cambridge, UK.

- Colby, J. & Zatman, L. *J. Biochem. J.* **128**, 1373-6 (1972).
- Anthony, C. & Zatman, L. *J. Biochem. J.* **104**, 960-969 (1967).
- Sperl, G. T., Forrest, H. S. & Gibson, D. T. *J. Bact.* **118**, 541-550 (1974).
- Duine, J. A., Frank, J. & Westerling, J. *Biochim. biophys. Acta* **524**, 277-287 (1978).
- Patel, R. N., Hou, C. T. & Felix, A. *J. Bact.* **133**, 641-9 (1978).
- Wadzinski, A. M. & Ribbons, D. W. *J. Bact.* **122**, 1364-74 (1975).
- Sperl, G. T. thesis, Univ. Texas (1973).
- Sperl, G. T., Forrest, H. S. & Hoare, D. S. *Bact. Proc.* 151 (1973).
- Dahl, J. S., Mehta, R. J. & Hoare, D. S. *J. Bact.* **109**, 437-445 (1972).
- Johnson, C. K. *ORTEP Rep.* ORNL-3794, (Oak Ridge National Laboratory, Tennessee, 1965).

matters arising

Variation with time of a Sun-weather effect

THE reported¹ influence of the solar and interplanetary magnetic sector structure on terrestrial atmospheric vorticity at the 500-mbar level during the winters of 1963–73 has been disputed by Williams and Gerety². We consider, however, that the response has been remarkably constant if an apparent decrease during the past few years of the intensity of tropospheric circulation is properly accounted for.

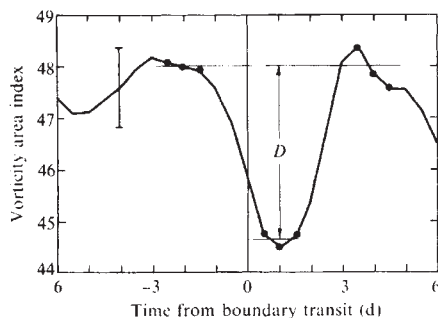


Fig. 1 A superimposed epoch analysis of the vorticity area index at 500 mbar about 162 times of interplanetary magnetic sector boundary transits during the winters in the interval 1 November 1963 to 31 March 1976 for which spacecraft observations of boundary transits are available. A typical error bar (twice the s.e.m.) is shown for the point at day -4.

First, Hines and Halevy³ concluded that "If solar sector boundary crossings are to be employed in the determination of key dates, their times of occurrence must be identified to at least an accuracy of ± 24 h on average for the analysis to be successful. This constraint is not serious when spacecraft data are employed for the determination of field reversals, but it can become significant when indirect determinations are made from ground-based geomagnetic measurements"^{4,5}.

Second, Hines and Halevy³ introduced the excursion x , defined as the difference between the maximum and minimum values of the vorticity area index (VAI) found in a 12-d interval centred on the time of boundary transit. The amplitude of the Sun-weather influence was small when the excursion was small and large when the excursion was large. In the past few years the observed excursions have been considerably smaller than in the previous years.

The interval during which spacecraft observations of the interplanetary magnetic field are available is the winter months from 1 November 1963 to 31

March 1976. The response of the VAI at 500 mbar to 162 sector boundary transits past the Earth during this spacecraft interval is shown in Fig. 1. The previously reported¹ minimum in VAI approximately one day after boundary transit is clear. A measure D of the depth of this minimum is defined as the average VAI near days -2 and +4 (w.r.t. day 0 as the boundary transit time) minus the average VAI near day 1. For each specified day the average VAI during three adjacent half-days is computed. In Fig. 1, $D = 3.30 \times 10^5 \text{ km}^2$.

D is found³ to increase as x increases; x_s divides the 162 boundary transits during the spacecraft interval into two equal groups, one with larger excursions and one with smaller. The Sun-weather effect is examined separately for each group.

Consider now the three-winter interval from 1 November 1963 to 31 March 1966. Figure 2 shows, for this interval (plotted at 1965), the average value of D associated with the group of boundary transits having excursions $> x_s$, and the average D for the group of boundary transits having excursions $< x_s$. The analysis is repeated one year at a time between 1 November 1975 and 31 March 1978 (plotted at 1977), the last for which data are available.

Between 1963 and the present shown in Fig. 2, the size of the Sun-weather associated with the group of boundaries having larger excursions is rather constant. For each interval shown in Fig. 2, D is larger for the group of boundaries associated with larger excursions than for the group of boundaries associated with smaller excursions, thus confirming in a much larger data set the result of Hines

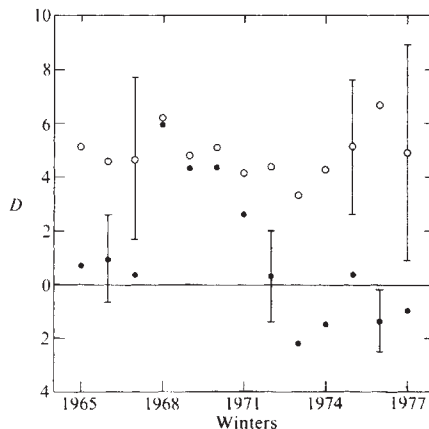


Fig. 2 D for the groups of boundary transits having larger excursions (○) and for the groups of boundary transits having smaller excursions (●). The results for the last winters should be considered preliminary until spacecraft observations of boundary transits are available. The total length of the error bar is twice the s.e.m.

and Halevy³. Except for the winters 1968–71, D associated with the group of boundary transits having smaller excursions is close to zero.

During the winters of the spacecraft interval shown in Fig. 3 the number of boundary transits whose excursions were $> x_s$ and the number of boundary transits with excursions $< x_s$ are approximately equal (note x_s was defined to make them equal during the entire spacecraft interval). In recent winters the magnitude of the excursions has declined considerably such that in the last interval only seven boundary transits had excursions $> x_s$, while 31 transits did not. If this decline is not an artefact of the meteorological data processing, an important change in the large-scale tropospheric circulation in the Northern Hemisphere has occurred in the past few years.

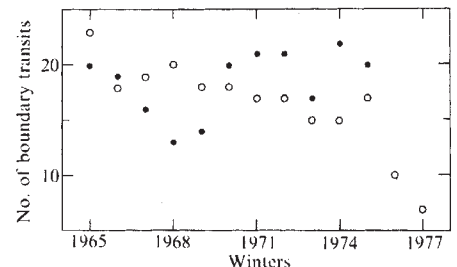


Fig. 3 The number of boundary transits in each three-winter interval for which the excursions are in the larger group (○) and in the smaller group (●). Note that in the last two intervals the number of boundary transits with larger excursions is considerably decreased.

In the winters before the spacecraft interval only indirect determinations of boundary transits made from ground-based geomagnetic measurements are available. Using these indirect boundary transits no consistent Sun-weather effect similar to that shown in Fig. 2 is found. It is quite possible that there was a variation in the Sun-weather effect in these earlier years, but the situation is complicated by the lack of spacecraft observations.

This work was partly supported by the office of Naval Research under Contract N00014-76-C-0207, by NASA grant NGR 05-020-559 and contract NAS5-24420, and by NSF grant ATM77-20580.

J. M. WILCOX
P. H. SCHERRER

*Institute for Plasma Research,
Stanford University,
Stanford, California 94305*

1. Wilcox, J. M., Svalgaard, L. & Scherrer, P. H. *J. Atmos. Sci.* **33**, 1114 (1976).
2. Williams, R. G. & Gerety, E. J. *Nature* **275**, 200 (1978).
3. Hines, C. O. & Halevy, I. *J. Atmos. Sci.* **34**, 382 (1977).
4. Berthelier, A. & Guerin, C. J. *geophys. Res.* **80**, 4390 (1975).
5. Svalgaard, L. *J. geophys. Res.* **80**, 2717 (1975).

WILLIAMS REPLIES—The biggest weakness of the study by Wilcox and Scherrer is immediately apparent. In subdividing the keydates according to their excursion, Wilcox and Scherrer have allowed themselves freedom to select the data *a posteriori*. Despite the work of Hines and Halevy¹ there is no known physical reason for selecting the data in this manner. In this context, it must be remembered that the data has already been carefully selected, for example the wintertime, 500 or 300 mbar VAI is chosen since it exhibits the best result. Unfortunately, in the absence of a physical understanding this is one of the few ways in which we can progress. Indeed, I have been working along similar lines².

A lesser weakness has been pointed out by Hines and Halevy¹. If one hypothesises that the 'signal' is the remnant of the 'noise' in the VAI dataset, then one might also expect the signal to increase as the noise increases. However, one would not expect the consistency claimed by Wilcox and Scherrer. A new point along these lines concerns the definition of the excursion. If the signal is real, then the strength of the signal can hardly depend on the meteorological noise after day 0 since then one would merely be correlating the signal with itself. Therefore, I would like to see the analysis repeated, this time selecting keydates according to the excursion before the start of the signal.

With regard to more minor points, I would like to see the results of the same analysis at 300 mbar; and how sensitive is Fig. 3 to a small change in x_s ?

If the signal is real and does depend on the intrinsic variability of the atmosphere, then this is an important result which ought to have strong implications for the search for a mechanism. In this case the study of Wilcox and Scherrer should be taken further to answer some important questions. For example is there a critical excursion below which no effect occurs? Is there a proportionality between the excursion and the amplitude of the signal? (The winters 1968–71, which could be interpreted as 1967–71 in view of the 3-y average employed and the '1967' and '1972' results, seem to suggest this is not so.) Can the winter–summer asymmetry of the result be explained in this manner, since one could expect the excursion to be less in summer?

The simplest interpretation of the VAI–SSB effect still seems to be statistical chance. This belief has been reinforced by an interpretation of this effect in terms of the tropospheric energetics in the wavenumber domain³.

R. GARETH WILLIAMS

Physics Department,
The University,
Southampton, UK

1. Hines, C. O. & Halevy, I. J. *atmos. Sci.* **34**, 382 (1977).
2. Williams, R. G., Donnegan, M. & Rycroft, M. J. *Nature* (submitted).
3. Williams, R. G. & Gerety, E. J. *J. atmos. Terr. Phys.* (in the press).

Anionic copolymerisation of episulphides with elemental sulphur

PENCZEK *et al.*¹ believe that the anionic copolymerisation of propylene sulphide (PS) with elemental sulphur gives the high-molecular weight polysulphide polymers $-\text{CH}_2-\text{CH}(\text{CH}_3)-\text{S}_x-$ (% S up to 90%, that is, $x = 12$). Nevertheless, PS mainly yields copolymers with $x \leq 2$ (refs 2, 3). We believe that this contradiction^{1–3} is caused by the presence of residual sulphur in the PS copolymers¹, which can be detected in the elementary analysis and does not show up in the NMR spectra. This assumption is based on our experimental data² and also confirmed by ¹³C NMR data¹ which indicate that PS copolymers with $x \leq 2$, and not polysulphide polymers with x up to 12 (ref. 1), are being formed. Indeed, polypropylene sulphide ($x = 1$) is characterised by the signals $\delta 20.7$ for CH_3 , $\delta 38.6$ for CH_2 and $\delta 41.0$ and $\delta 41.2$ for CH^4 , whereas polypropylene disulphide ($x = 2$)⁵ is characterised by the signals $\delta 19.05$ for CH_3 , $\delta 44.6$ and $\delta 45.0$ for CH_2 and $\delta 46.0$ and $\delta 46.4$ p.p.m. for CH^6 . The spectra of PS copolymers with $2 > x > 1$ also have the signals of conjunctive units $-\text{SS}-\text{CH}_2-\text{CH}(\text{CH}_3)-\text{S}-\text{CH}_2-\text{CH}(\text{CH}_3)-\text{SS}-$ at $\delta 39.9$ and $\delta 39.5$ for CH , $\delta 37.4$ for CH_2 and $\delta 19.4$ p.p.m. for CH_3 (ref. 6). Therefore the increase in the intensity of signals 45.8, 45.6, 39.5, 39.1 and 19.4 p.p.m. with the increasing sulphur content in PS copolymers¹ are due to the increase in the amount of disulphide bonds ($x = 2$), and not polysulphide bonds ($x > 2$) as other signals in the spectra are absent (for example, CH_3 signal of trisulphide units should appear at $\delta 18.1$ p.p.m.⁶). It follows that the apparent values of $x = 2$ and 5 in the spectra of PS copolymers¹ conform to $x \leq 1.5$ and 1.7, respectively. Like the ¹³C NMR data^{1,2}, ¹H NMR spectra of PS copolymers² correspond to that of PS polymers with a varying ratio of mono- and disulphide units⁶.

We have shown previously^{2,3} that the probability of a polysulphide polymer formation increases as the nucleophilic reactivity of episulphides decreases. Very reactive ethylenesulphide is predominantly converted into polyethylene sulphide ($x = 1$). Polysulphide polymers ($x > 2$) are mainly produced from a less reactive cyclohexene sulphide, isobutylene sulphide and trimethylethylenesulphide. The latter forms the copolymers with x up to 7 whose ¹H NMR spectra show a gradual downfield shift of proton signals as x is increased in the case of more reactive PS. The probability of homopolymerisation and formation of stable disulphide units increases as unstable polysulphide chains must split at hard reaction conditions¹. The stereochemical effect of substituents

and more mild copolymerisation conditions² apparently decrease the homopolymerisation rate of substituted episulphides and thereby increase the probability of their copolymerisation with sulphur as well as stabilise and restrict the participation of polysulphide chains (at least for $x \leq 5$) in nucleophilic splitting reactions because of the sterically hindered growing thiolate anions. Thus, despite these limitations, which can be overcome by using active forms of sulphur, inorganic polysulphides or UV irradiation², this method offers broad possibilities in the synthesis of polysulphide polymers^{2,3}.

A. D. ALIEV
ZH. ZHUMABAEV
B. A. KRENTSEL

USSR Academy of Sciences,
A. V. Topchiev Institute of
Petrochemical Synthesis,
Leninsky prospekt, 29
Moscow, USSR

1. Penczek, S., Slazak, R. & Duda A. *Nature* **273**, 738–739 (1978).
2. Aliev, A. D., Solomatina, I. P., Zhumabaev, Zh., Alieva, S. L. & Krentsel, B. A. Auth. cert. USSR N516711 (1976); *Byull. Izobr.* **21**, 91 (1976). *Referat. Zh. Khim.* **11**, C399 II (1978).
3. Aliev, A. D., Solomatina, I. P., Zhumabaev, Zh. & Krentsel, B. A. XIV Scientific Session on Chemistry and Technology of Organic Sulfur Compounds and Sulfurous Oils (Abstr.) 224–226 (1976). *Referat. Zh. Khim.* **15**, C229 (1976).
4. Boileau, S., Sheradame, H., Guerin, P. & Sigwalt, P. *J. chim. Phys.* **69**, 1420–1424 (1972).
5. Aliev, A. D., Solomatina, I. P., Koshevnik, A. Yu., Zhumabaev, Zh. & Krentsel, B. A. *Vysokomolek. soedin.* **A19**, 173–180 (1977).
6. Aliev, A. D., Alieva, S. L. & Krentsel, B. A. *Vysokomolek. soedin.* (in the press).

PENCZEK ET AL. REPLY—Aliev *et al.* have confirmed a general principle of anionic copolymerisability of elemental sulphur¹, including copolymers with various cyclic sulphides as described in our patent². However, Aliev *et al.* assume that copolymers of propylene sulphide with elemental sulphur could have merely be solid solutions of polysulphide polymers $-\text{CH}_2-\text{CH}(\text{CH}_3)-\text{S}_x-$ with an average $\bar{x} < 2$ and elemental sulphur. This was based on their failure to prepare in their system copolymers of propylene sulphide with higher sulphur content, and on their interpretation of the ¹³C-NMR spectra. However, great care must be taken with the interpretation of the NMR spectra of polysulphides, because we observed recently that the upfield signals can be caused by the presence of the low-molecular weight cyclic polysulphides. These products were removed from our polymers by an extensive extraction with boiling methyl alcohol (unreacted sulphur was removed in the same way)³.

Had Aliev *et al.* attempted to repeat our experiments, leading to the true copolymers with high content of the built in sulphur, it would have become clear that, for example, a 50:50 mixture of polypropylene sulphide and elemental

sulphur (this mixture gives an apparent average $\bar{x} = 4.3$) dissolved in a common solvent (such as, CS_2) gives on evaporation of the solvent an opaque, heterogeneous mixture, from which good crystals of sulphur crystallise out. This behaviour sharply contrasts with that of copolymers with $\bar{x}$ even as high as $\bar{x} = 8$ (85.9% of sulphur) which give transparent, non-opaque films from solutions.

Another equally powerful proof of the absence of nonreacted sulphur in copolymers comes from our recent studies of these and related copolymers³ by laser Raman method (Cary 82). Elemental sulphur shows two strong peaks in benzene solution, at 217 cm^{-1} and 475 cm^{-1} , well separated from other signals in the intentionally prepared mixtures of polysulphides with elemental sulphur. These peaks are absent in copolymers of sulphur with propylene sulphide, prepared as previously described¹.

STANISŁAW PENCZEK
ROMUALD ŚLĄZAK
ANDRZEJ DUDA

*Polish Academy of Sciences,
Centre of Molecular and
Macromolecular Studies,
90-362 Łódź, Boczna 5, Poland*

1. Penczek, S., Ślązak, R. & Duda, A. *Nature* **273**, 738–739 (1978).
2. Polish Patent, 78934 (1976).
3. Duda, A. & Penczek, S. *Makromolec. Chem.* (in the press).

Short-range structure of amorphous alloys

WANG'S problem of the structures of amorphous intertransition metals¹ is two-fold. First, what is the nature of the short-range chemical ordering? It is generally considered that it is probably related to the ordering in the corresponding crystal, although quantitative evidence, based largely on neutron scattering^{2,3} and EXAFS⁴ data, is still inadequate. There is no guarantee that the particular clustering of type A atoms about type B observed in the crystal will still be consistent with stability when the 'surrounding matrix' is disordered rather than crystalline, especially where the chemical forces are relatively weak. At the very weak extreme, the evidence is strongly against either crystalline or icosahedral (polytetrahedral) short-range ordering in ideal single component glasses^{5–7}.

The second, more interesting problem of structural modelling is: how can we construct an extended, dense, non-crystalline model which is both stable and consistent with the short-range chemical ordering (refs 5, 6)? We need to consider, (1) the topology of the connectedness between like and unlike component atoms, and (2) the packing constraints consistent with dense, extended non-crystalline systems. Neither of these is

understood at all well, but we do know the problems are of a very different nature in two and three dimensions. For example, the triangulation of dense, disordered two-dimensional assemblies is straightforward and can be done with only minor distortions, as Wang shows empirically. The equivalent tetrahedral division of three-dimensional assemblies without the presence of additional octahedra or other more complex polyhedra is very much more difficult, and probably impossible. We cannot pack polytetrahedral units together without serious boundary misfits which cannot be accommodated by small distortions⁵. Second, the connective topology problems in two dimensions are also much simpler than the equivalent three-dimensional problem. Thus, whereas the misfits at polyhedron boundaries and the atomic disorder consequent on non-crystallinity, are relatively small in Wang's two-dimensional model, we learn very little about the equivalent problem in three dimensions, which is the space in which real amorphous alloys are found. It is unfortunate, though understandable, that the discussion is of a two-dimensional model. We cannot extrapolate into three dimensions, where the connective topology and packing constraints (Bernal's 'statistical geometry' (ref. 8)) are so poorly understood.

Wang comments that dense random packing of hard spheres (DRPHS) models fail to explain experimental data on these systems. This assumes an over-narrow interpretation of such models. Unfortunately, the specific structure¹⁰ of the hard-sphere model, rather than the underlying concept of a dense, non-crystalline organisation consistent with the interatomic interactions⁵, has often been applied incorrectly to binary alloys⁹. The value of the original model¹⁰ was largely conceptual, focusing attention away from crystallites to models that could be discussed within an inherently non-crystalline framework. To apply such a simple, idealised model unaltered to a variety of binary systems is clearly incorrect⁵: even weak chemical ordering and slight softening of potential functions are known to result in a variety of non-crystalline structures, all conceptually related to DRPHS by their non-crystalline nature, but different from each other in detailed local structure⁶. Moreover, most computer-built so-called DRPHS models are not dense random packings; because of problems of accessing the high-density states by computer construction^{5,6}, most models are of a relatively low density^{5,11}, and hence will fail to reproduce the necessary packing constraints in the real system. The failure that Wang quotes¹ is not of the DRP concept, but of its inadequate realisation in a particular system.

The problem of binary amorphous alloy structure is how to feed in the chemical ordering when constructing a dense model so that the resultant model is consistent

with both the ordering and the packing constraints consequent upon the high experimental density. This is far from easy, but it needs tackling quantitatively in three dimensions. The work of Gaskell¹², Boudreaux¹³ and von Heimendahl¹⁴ are useful attempts to solve these difficulties.

J. L. FINNEY

*Department of Crystallography,
Birkbeck College, University of London,
Malet Street, London WC1, UK*

1. Wang, R. *Nature* **278**, 700–704 (1979).
2. Sadoc, J.-F. & Dixmier, J. *Mater. Sci. Engng* **23**, 187–192 (1976).
3. Kudo, T. *et al.* *J. phys. Soc. Japan* **45**, 1773–1774 (1978).
4. Hayes, J. M., Allen, J. W., Tauc, J., Giessen, B. C. & Hauser, J. *J. Phys. Rev. Lett.* **40**, 1282–1285 (1978).
5. Finney, J. L. *Nature* **266**, 309–314 (1977).
6. Finney, J. L. in *Diffraction Studies of Non-Crystalline Substances* (eds Hargittai, I. & Orville-Thomas, W. J.) (Akademiai Kiado, Budapest, in the press).
7. Barker, J. A., Hoare, M. R. & Finney, J. L. *Nature* **257**, 120–122 (1975).
8. Bernal, J. D. *Proc. R. Soc. A* **280**, 299–322 (1964).
9. Cargill, G. S. *J. appl. Phys.* **41**, 12–29 (1970).
10. Finney, J. L. *Proc. R. Soc. A* **319**, 479–493 (1970).
11. Cargill, G. S. & Kirkpatrick, S. in *Structure and Excitations of Amorphous Solids*, 339 (American Institute of Physics, 1976).
12. Gaskell, P. H. *Nature* **276**, 484–485 (1978).
13. Boudreaux, D. S. & Gregor, J. M. *J. appl. Phys.* **48**, 5057–5061 (1977).
14. von Heimendahl, L. (in preparation).

WANG REPLIES—Like the original DRP concept, the proposed short-range structure of amorphous intertransition metal alloys is largely conceptual. This structural concept can be verified by experiments and structural modelling, but at present, the experimental evidence, based on limited scattering and EXAFS data, is inadequate to prove or disprove the proposed concept. Concerning structural modelling, Finney's past work can be summarised as "one cannot pack idealised polytetrahedra or spherically-symmetrical atoms to reach a high packing density" (ref. 1). New structural modelling techniques should be developed using extremely soft potentials to account for the mixing of irregular tetrahedra and the asymmetrically shaped atoms as suggested in the structural concept. Furthermore, if one analyses Samson's nine modes² for connecting truncated tetrahedra in crystalline compounds, one can see that the boundary problem in three dimensions may not be as hopeless as Finney suggests.

The main question is: to what extent is crystalline-like short-range order preserved in the real amorphous structure? This is difficult to answer because both composition and preparation method can influence the short-range structure. The real structure should be a mixture of metallic and covalent bondings with localised coordination numbers varying from 8 to 16. So far, the experiments and structural modelling do not uniquely define a real 'amorphous' structure. As Rustum Roy said, "for non-crystalline phases in general, it is only our imagination that limits our models . . . of homogeneous and heterogeneous structures." I

hope, through this structural concept, that we can systematise our knowledge about the formation, stability and unique properties of the amorphous intertransition metal alloys and relate these properties to those of the crystalline counterparts.

R. WANG

Materials Department,
Pacific Northwest Laboratories,
Richland, Washington 99352

1. Finney, J. L. *Nature* **266**, 309–314 (1977).

2. Samson, S. in *Developments in the Structural Chemistry of Alloy Phases* (ed. Giessen, B. C.) 65–106 (Plenum, New York, 1969).

Circular hydrogen bonds

SAENGER¹ rightly stresses both cooperativity and large deviations from tetrahedrality in water-involved hydrogen-bonding networks. Similar lattice unconstrained structures were observed previously when examining the hydration of biological macromolecules, especially proteins^{2–4}, although the α -cyclodextrin data examined by Saenger is of significantly better quality, and therefore more quantitative.

However, although I agree strongly that cooperative effects are relevant in OH-group hydrogen bonding^{3–7}, the existence of rings—be they homodromic or heterodromic—does not of itself conclusively indicate that cooperative effects are operating. Given even a purely pair-additive potential with an average coordination of more than two, we would expect ring structures to be a topological consequence. Current ideas on the structures of trigonal⁸ (for example, B_2O_3) and tetrahedral⁹ (for example SiO_2) glasses necessarily involve rings without any known cooperativity. As in aqueous systems, simulation calculations yield extensive ring structures using purely pair-additive potentials¹⁰. The wide glass-forming ability of the silicate glasses, and the many crystalline forms of silicates, underline the structural versatility of tetrahedral coordination; the ring structures are a consequence of the (non-cooperative) interaction potential and the need to fill space within the constraints of the potential.

The α -cyclodextrin crystal data are valuable for the quantification of deviations from ideal tetrahedrality, an effect previously noted with respect to poorer quality protein data^{4,11}. The conventional wisdom of tetrahedrality in water is based on (1) the geometry of sp^3 hybridisation; and (2) the structures of the ices. The ice structures are, however, subject to the lattice constraint; extrapolation to small disordered clusters of molecules is dangerous. Concerning (1), most recent large basis set *ab initio* calculations on the water molecules¹² show the lone pairs

poorly separated; the deviation from tetrahedrality is large. This is backed up by neutron diffraction studies of hydrates, which show the directional control of the lone pairs to be relatively poor^{7,13}. Strong tetrahedrality is also inconsistent with recent neutron scattering data on liquid water (J. C. Dore and I. P. Gibson, in preparation; refs 14, 15); pair correlation functions predicted by non-cooperative tetrahedral models of the water molecule show far more short range tetrahedral structure than is observed¹⁵. In contrast, a cooperative potential which makes no such geometrical assumptions^{5,6} is in agreement with the data (P. Barnes, J. L. F., J. E. Quinn and J. D. Nicholas, in preparation).

Cooperativity and imperfect tetrahedrality are indeed important aspects of aqueous systems which are normally ignored^{3–7}. Saenger's data give quantitative support to imperfect tetrahedrality, but does not of itself conclusively indicate that cooperativity is operating.

J. L. FINNEY

Department of Crystallography,
Birkbeck College, University of London,
Malet Street, London, WC1, UK

1. Saenger, W. *Nature* **279**, 343–344 (1979).

2. Berendsen, H. J. C. in *Water: A Comprehensive Treatise*, Vol. 5 (ed. Franks, F.) 293–330 (Plenum, New York, 1975).

3. Finney, J. L. *Phil. Trans. R. Soc. B* **278**, 3–32 (1977).

4. Finney, J. L. in *Water: A Comprehensive Treatise* Vol. 6 (ed. Franks, F.) (Plenum, New York, in the press).

5. Barnes, P. in *Progress in Liquid Physics* (ed. Croxton, C. A.) 381–428 (Wiley, Chichester, 1978).

6. Barnes, P., Finney, J. L., Nicholas, J. D., Quinn, J. & Westerman, A. V. *Acta crystallogr. A* **34**, Suppl., S200 (1978).

7. Finney, J. L., *Farad. Disc. chem. Soc.* **66** (in the press).

8. Mozzi, R. L. & Warren, B. E. *J. appl. Crystallogr.* **3**, 251–257 (1970).

9. Bell, R. J. & Dean, P. *Nature* **212**, 1354–1356 (1966).

10. Rahman, A. & Stillinger, F. H. *J. Am. chem. Soc.* **95**, 7943–7948 (1973).

11. Finney, J. L. *J. molec. Biol.* **119**, 415–441 (1978).

12. Dierksen, G. H. F. *Theor. chim. Acta (Berl.)* **21**, 335–367 (1971).

13. Olovsson, I. & Jonsson, P.-G. in *The Hydrogen Bond* Vol. 2 (eds Schuster, P., Zundel, G. & Sandorfy, C.) 393–456 (North-Holland, Amsterdam, 1976).

14. Gibson, I. P. thesis, Univ. Kent (1978).

15. Dore, J. C. *Farad. Disc. chem. Soc.* **66**, (in the press).

SAENGER REPLIES—From a topological point of view, it is certainly justified to assume that water molecules and hydroxyl groups could form circular H-bonded structures without invoking the cooperative effect. The quadrupole functionality of water in combination with purely pair-additive interaction potentials could be sufficient to arrange water molecules into circular aggregates, similar to the formation of circles in SiO_2 or B_2O_3 glasses, as pointed out by Finney. However, we should be aware that in $Si-O-Si$ and $B-O-B$ bonds the directions of the bonds (from donor to acceptor) are predetermined because O is the electron acceptor and Si or B are the donors. It is one of the special properties of H-bonds between hydroxyl groups $O-H\cdots O$ that both oxygen atoms can function either as donor or as acceptor. This duality is

clearly resolved if H-atoms are located and, because the direction of the H-bonds is then known, the existence or absence of cooperativity between several such H-bonds is directly evident.

In the absence of a cooperative effect, we should expect near-random directions of $O-H\cdots O$ hydrogen bonds, be they in 'endless' chains or in circular structures (which then would be heterodromic). Because the neutron data on α -cyclodextrin hexahydrate have enabled us to locate all H-atoms unambiguously, the directions of all the H-bonds are known. We find that in the two chain-like structures as well as in the homo- and anti-dromic circles all the H-bonds are unidirectional, a strong indication for the existence of the cooperative effect. Obviously, both topological and cooperative effects favour and stabilise circular H-bonded structures.

Bogdan Lesyng (Warsaw University) has performed PCILO calculations on chain-like and circular H-bonds (in preparation). He compared H-bonding energies of these structures with those of the respective pairwise, isolated $O-H\cdots O$ interactions. The influence of the cooperative effect is marginal, about 1% of the total H-bonding energy, in the two antidromic circles, but 7.3% in the homodromic circle and 9.8% in the chain-like structure. The data suggest that cooperativity does exist in these H-bonded structures but its magnitude depends strongly on the arrangement of the H-bonds, one of the bonds in the antidromic circle III even being bifurcated. Another interesting result of these calculations concerns the dipole moments, 2.73 D for the homodromic circle but ≥ 6 D for the antidromic circles and chain-like structure, suggesting that in water clusters, homodromic circles should be the preferred species.

WOLFRAM SAENGER

Abteilung Chemie,
Max-Planck-Institut
für experimentelle Medizin,
D-3400 Göttingen, FRG

Matters Arising

Matters Arising is meant as a vehicle for comment and discussion about papers that appear in *Nature*. The originator of a Matters Arising contribution should initially send his manuscript to the author of the original paper and both parties should, wherever possible, agree on what is to be submitted. Neither contribution nor reply (if one is necessary) should be longer than 300 words and the briefest of replies, to the effect that a point is taken, should be considered.

reviews

Participation as an ideology

Eric Ashby

Technological Decisions and Democracy: European Experiments in Public Participation. By Dorothy Nelkin. Pp. 111. (Sage Publications: Beverley Hills and London, UK, 1978.) £6.

If you have an hour or two to wait in Chicago airport between flights you are likely to be approached by earnest young people asking you to sign a protest against the use of nuclear power in America, following the incident at Harrisburg. It is ironic, in the light of the recent air tragedy there, that you would not be asked to sign a protest against the use of passenger aircraft. This is a symptom of the basic irrationality in public opinion about the applications of technology in America. It is a cause for concern among policy-makers. They know that in a pluralistic democracy the pressure for participation cannot be resisted and they know that satisfactory patterns for participation have not yet been worked out. It is clear that in all western democracies people are no longer content to leave decisions to experts; but the alternative is disquieting because the decisions cannot possibly be made without experts. If popular views about risks, costs and benefits coincided with the calculations of statisticians there would be some hope for rational discourse. But they don't. And the dilemma is compounded by the disturbing fact that protests against nuclear power, the siting of airports, and even research into genetic recombination easily slide into a much deeper protest against the concentrations of power inevitable in representative democracy. Given that participation has come to stay there is an urgent need for innovations to incorporate participation into our political system in such a way that it strengthens and does not clog up the machinery of government. The questions in need of answers are: Who should participate? How can those who participate be well enough informed to make a useful contribution? What pattern of participation (an adversary system or an inquisitorial system) would be more effective?

Dorothy Nelkin, in the Program on Science, Technology and Society at Cornell, has published several scholarly case-histories on this problem. In this brief essay she has turned to enquire how participation works in three European countries — Sweden, Holland and Austria — where decisions can be made in a much simpler political context than exists in the

United States. Sweden is perhaps the most cohesive community in Europe, with a central government accustomed to a tradition of "consensus maintained by compromise". Holland is divided by religious and regional loyalties which may coalesce over technological and scientific issues, or may use such issues as ammunition for political conflicts. Austria is a federal republic which has recently succeeded in breaking away from the strong teutonic traditions of authoritarianism. All three countries have reacted to popular demands for participation in decisions about planning, the generation of energy, and science policy. In each country, public interest was ignited by incidents which were given saturation treatment in the media. In Sweden there were incidents like mercury poisoning, acid rain, and the energy policy for a nation without coal or oil but with a supply of uranium. In Holland action groups which had opposed the location of blast furnaces in Rotterdam, the location of an airport, and the building of a dam across the Oosterschelde combined to form the Foundation for Nature and the Environment, with some 100,000 members. In Austria, opposition to nuclear power culminated in a referendum last year in which some 64% of the electorate voted and which by a narrow majority rejected a newly built power station already charged with nuclear fuel.

Professor Nelkin describes how each of these countries has devised experiments in participation over issues in nuclear energy and science policy. The most significant experiment is the one devised in Sweden. For many years there have been study circles there, run by trade unions, religious groups, and other social organisations. In May 1974 the Riksdag approved a proposal to sponsor energy studies by these groups. Twelve thousand study circles, involving some 80,000 people, were subsidised to the tune of about £330,000. This did not produce any dramatic enlightenment among the Swedish people but it did mean that the government's Energy Policy Bill was more widely read, and — an unexpected result — the very complexity of the issues, when these were brought to public attention, increased the number of 'don't knows' in the pro- and anti-nuclear power controversy. In Holland the most interesting experiment is the organisation of courses in the University of Amsterdam to teach scientists how to popularise research findings, and in Utrecht University,

courses for working journalists to broaden their knowledge of science and technology.

Professor Nelkin's essay is written with the clarity and precision characteristic of her earlier writings. Apart from its value as a summary of information not readily available elsewhere, it will stimulate the American and British reader to reflect on the arrangements for participation in their own countries. Political conflict and ambiguity, she writes, "are basic realities of technological decisions". To expect sweetness and light over Windscale, a third London airport, or coal mining in the Vale of Belvoir, is a delusion. It is a delusion also to expect technical decisions of this kind to remain technical: they are bound to become political. It is for this very reason that the public (in my view) have a right to participate. What, then, are reasonable conditions for participation?

First, if the authorities take the trouble (and it does involve trouble) to organise participation, then the participants ought to take the trouble to do their homework. Some preliminary education which is not just propaganda is essential. This means access to information and to the advice of experts who are not *parti pris* to the issue. Second, participation will not be acceptable if it is mere token consultation; there must be unambiguous assurances that the views of participants will be taken into account. (Enquiries under the Town and Country Planning Act do give the British public such assurances: witness the many enquiries in which the Minister has come down on the side of conservationists, from the siting of the Cambridge biochemistry department to the siting of a power station near Nottingham). Third, it is unrealistic to suppose that a majority of the public will take an intelligent interest in any major technological issue. There will be a spectrum of opinion, with persons of impenetrable prejudice at each end of the spectrum and a mass of apathetic people in the middle. But between these three groups are the people who matter. They already hold an opinion but they are willing to subject their view to further argument. They are sufficiently concerned to give time to attending meetings, reading articles, and joining pressure groups. These are the people who should be drawn into active participation. Their views ought to be taken seriously. It will not do to dismiss their opinions as coming from an élite minority, unrepresentative of the mass of opinion; for unless the mass takes the

trouble to express an opinion, it has to be assumed that it has no opinion. Nor will it do to say that these groups of concerned people do not understand the highly technical issues; it is the duty of the authority (the Central Electricity Generating Board, or the Ministry of Transport, or whoever proposes the project under discussion) to set out the issues in a way which can be understood by laymen.

We have an opportunity for an experiment in participation in England. The proposal to mine coal in the Vale of Belvoir is going to arouse savage opposition. The issues are highly complex. Would it be possible for the adjoining universities (Nottingham, Leicester, and Loughborough) to arrange extra-mural

courses about the issue, conducted by experts in geology, energy technology, and economics who would not be regarded as in the pay of the Coal Board? And — but perhaps this is going too far — would it be possible for witnesses holding a class certificate from one of these extra-mural courses to be given preference in the public enquiry which will doubtless be held? A procedure like this might not lead to a wiser decision (though it might do that too), but it would be likely to increase confidence and sympathy for the decision. And in a pluralistic democracy the process is as important as the end product. □

Lord Ashby is Chancellor of Queen's University, Belfast, and a Fellow of Clare College, Cambridge, UK.

Bumblebee economics

Bumblebee Economics. By Bernd Heinrich. Pp. 159. (Harvard University Press: Cambridge, Massachusetts, and London, UK, 1979.) \$21.90; £11.40.

FIVE books dealing entirely with bumblebees have been published previously, in 1912, 1934, 1959, 1975 and 1978, and three other books, published in 1970, 1971 and 1974, have substantial parts devoted to bumblebees. The justification for producing another book on bumblebees would need to rest on significant advances in the subject or on a new approach. Certainly Bernd Heinrich gives us a refreshingly different and topical approach.

The central theme running through his book is that of energy economics, and most of the activities of individual bumblebees and of bumblebee colonies are examined in the light of profit and loss of energy resources. Bumblebees are particularly suited for such an examination. They live in temperate climates and often need to survive low temperatures. To ensure that the susceptible brood does not become chilled, the adults produce heat by metabolic activity and surround the comb with insulating material to retain it.

However, the ultimate object of a bumblebee colony is to convert the nectar and pollen collected by its foragers into sexually developed adults that mate and reproduce. Calculating the energy balances and the effects of likely changes in the collection, flow and expenditure of resources is shown to be fascinating but complex. For example, bumblebees are well known for their ability to fly and forage at relatively low temperatures and in adverse weather conditions, but the success of the currently in vogue phrase, 'foraging strategy', will depend on the energy they need to expend before flight is possible, the

amount of fuel (nectar) they consume to reach the foraging area, whether sufficient nectar is being produced, and whether it contains a sufficient concentration of sugars; both these latter factors are also controlled by environmental temperature and relative humidity. It is easy to envisage circumstances in which a forager would use more sugar than it actually collected. Furthermore foraging in poor weather conditions can be hazardous as well as arduous, and so diminish longevity and the energy resources of the bee's colony. Even in favoured foraging conditions several trips are needed to produce sufficient nectar and pollen to rear one worker bee, and Heinrich calculates that to rear one queen requires the nectar and pollen loads collected by a worker during two days of continuous foraging.

We are still lacking information on the precise mechanisms that instigate queen production. Increase in colony size and decrease in brood/worker ratio release a greater proportion of bees for foraging so the collection of sufficient forage for queen production should become possible. However, larger colonies need larger foraging areas to support them, so

individual foragers will need to forage further afield with diminished efficiency. As a colony becomes larger so does the average size of the worker so the larger forager and its greater load bearing capacity may more than compensate for an increased foraging area!

Heinrich also discusses how bumblebees interrelate with other animals and plants in their environment, and stresses especially the effect of bumblebee pollination on the production of fruits and seeds, that in turn provide the diet of much wildlife. In contrast to most insects that forage on plants, visits by bees actually increase the likelihood that the plants will reproduce, and so provide additional forage and energy requirements for succeeding generations of bees.

This book is written for the non-specialist; although easy to read the style is sometimes too colloquial for my taste. I see no point in calling empty cocoons "silken cradles" or to liken the adult bee to a "miniature helicopter" or "a small, furry, orange, black and yellow object". But few can avoid the temptation to try to wax poetic about the bumblebee.

Quite rightly, Heinrich pays particular attention to his own research and its implications. Although at times the reader is too frequently reminded to regard the bumblebee in terms of energy balance and allocation of resources, the central theme proves a useful one and serves the double purpose of binding the work into a coherent whole and allowing the author to return easily to his main line of thought after frequent digressions. There are two short appendices, on how to rear bumblebees and on the bumblebee species of North America, and a list of selected references.

This is a good book, packed with information. I recommend it as essential reading to those interested in social insects and hope it will find the wider readership it deserves.

John B. Free

John B. Free is at Rothamsted Experimental Station, Harpenden, UK.

Fluid compositions in the Earth's crust

Fluids in the Earth's Crust: Their Significance in Metamorphic, Tectonic and Chemical Transport Processes. By W. S. Fyfe, N. J. Price and A. B. Thompson. Pp. 372. (Elsevier Scientific: Amsterdam and New York, 1978.) \$49.75; Dfl. 125.

THE authors in the Preface of this book say that it is an attempt to link metamorphic processes and rock mechanics, to provide a text unifying the fields of metamorphic petrology and structural geology. In one sense, they have

admirably accomplished this task. The success of the book is that from the varied viewpoints and experiences of the authors, it is a synthesis of a portion of these complex and diverse fields. The general theme is the link between fluid compositions in the crust as controlled by equilibrium metamorphic mineral assemblages and kinetics, mass transport during metamorphic processes, and tectonics. The role of fluids in crustal processes is strongly emphasised. However, as the authors state, this text is not traditional or classic or all-encompassing.

Chapter 1 sets the stage, providing the focus for the link between fluid motion, geochemical and tectonic processes. It is a

short chapter, only 18 pages. In this, the authors point out that the question that concerns them in the book is how fluid flow is focussed, which is the question at the heart of economic geology.

Chapter 2 describes the chemistry of natural fluids and the characteristics and role of water in transport processes. It is essentially a brief review of the chemistry of surface waters and subsurface fluids, including pore waters, oil-field brines, metamorphic and magmatic fluids, and fluid inclusions.

Chapter 3 is brief and deals with the volatiles bound in minerals. In Chapter 4, the solubilities of major rock-forming minerals are discussed in terms of P-T-X diagrams. The concept that fluid compositions are determined by mineral solubilities modified by kinetic and residence time considerations is initially developed. The conclusion is drawn that natural fluid compositions can be adequately described in the model system $\text{NaCl-CO}_2\text{-H}_2\text{O}$.

Some elementary theory governing metamorphic mineral reaction rates is presented in chapter 5, and applied to solution rates of SiO_2 , CaCO_3 , and alkali feldspars. Rates of mineral nucleation and growth and diffusion are considered from a theoretical and empirical standpoint. Finally, rates and mechanisms of metamorphic reactions are discussed.

Chapter 6 deals with the temperatures and pressures of metamorphism and metamorphic facies. It emphasises how, where and when water is released during metamorphism and the composition of metamorphic fluids.

Rock-dominated and fluid-dominated metamorphism are discussed in Chapter 7. The behaviour of H_2O and CO_2 , O_2 and H_2 , S and SO_2 , halogens, and the behaviour of fluids during partial fusion, are emphasised. The role of fluid pressure and its relationship to P_{total} and $P_{\text{H}_2\text{O}}$ are critically assessed.

In chapter 8, there is a sharp change in direction. The subject is rock deformation and rock mechanics, particularly the experimental aspects. The strength of rocks is discussed in terms of the environmental parameters of confining pressure, temperature, water and fluid content and pressure, and time. The authors conclude that earlier high-strength data on rocks were incorrect. Unfortunately, this chapter omits any account of recent experimental work on

the weakening effect that the presence of water has on the creep strength of rock.

Using geological evidence, the authors attempt in chapter 9 to quantify the crustal conditions of temperature, confining pressure, pore fluid pressure, differential stress, and strain rate that control rock deformation. This is a very nice review of the 'state of the art'.

The concepts of permeability, hydraulic fracture, and elasticity are discussed in chapter 10, in preparation for the following two chapters. The treatment of fracture mechanics that is applied to hydraulic fracture is at an elementary level.

In chapters 11 and 12, controls on the escape of water during compaction, deformation and metamorphism, and igneous and salt diapirs and diapirism respectively, are considered. Very recent work on transport of magmas is not described in these chapters. In chapter 13, the role of fluids in tectonic and

metamorphic processes is synthesised and regions of large-scale crustal mass transport delineated. The authors certainly succeed in showing that the dynamics of a wet Earth is quite different from that of a dry Earth. (It should be pointed out that the conversion factor for viscosity from c.g.s. units to S.I. units given in the book is wrong.)

We found this book to be a pleasure to read, perhaps because it dealt principally with fields only somewhat allied to ours. It is well written and synthesises in a single package 'pieces of the puzzle' involving crustal fluids, and metamorphic and tectonic processes. We recommend it.

**Fred T. Mackenzie
J. Weertman**

Fred T. Mackenzie is Professor of Geological Science and J. Weertman is Professor of Materials Science and Engineering and Professor of Geophysics at Northwestern University, Evanston, Illinois.

Applied geology

Industrial Geology. Edited by J. L. Knill. Pp. 344. (Oxford University Press: Oxford, 1978.) £9.50.

THE most important aspect of this book is that it will bring to the attention of both teachers of geology and students alike, the industrial applications of the subject, of which many of them will have little practical experience or knowledge. Such a text would seem to be long overdue.

In fact the book arose from a series of lectures, given to students at the Department of Geology and Mineralogy at Oxford, by various industrial speakers whose contributions have now been put together to form this volume. Universities are not in a position — nor is it their function — to produce applied geologists, yet somehow undergraduates and especially postgraduates must be exposed to the practical aspects of the subject. The philosophy behind this concept is very clearly put down in Professor Vincent's foreword, and the first three paragraphs are recommended to everybody using this book. The artificial barriers between academic geology and applied geology as practised in industry are indeed attitudes of mind.

Almost every facet of industrial activity (but not Government) in which geologists are involved receives a chapter and the more important ones are given two; but the main aim is clearly to provide students with an introduction to the particular applications of geology in the mining, oil and civil engineering industries.

In the energy field, coal, oil and natural gas, the geologist has prime functions in

both in the exploration and production phases, and the chapters point out the varied nature of the geologists role and the related disciplines with which he must become familiar, for example, geophysics reservoir, engineering economics.

In the minerals industry the geologist is no less important but here perhaps relies less on the seismic method and more on mapping and regional petrogenesis. Three chapters are included and these are concerned primarily with exploration and metallurgy in the UK.

The extractive industries and the cement industry are then considered and these are followed in turn by descriptions of groundwater and the construction industry, geological hazards and conservation. It would be too lengthy a process to describe them in detail but even if the reader were expert in one field he would probably learn a great deal from reading the other chapters. So much more is to be gleaned, therefore, by the student coming fresh to the application of his subject.

Although not specifically intended as a careers guide, it will inevitably be used as one, not least to answer the question "what would I expect to do", but perhaps, more importantly, to show the student what particular facets (course options) of the subject will be best suited to the particular field eventually chosen.

The book deserves to be widely read not just by students and teachers, although they will have most to gain, but by all geologists.

J. R. V. Brooks

J. R. V. Brooks is Senior Geologist in the Petroleum Engineering Division of the Department of Energy of Her Majesty's Government.

● *The Pursuit of Nature: Informal Essays on the History of Physiology* (for the review, see *Nature*, 270, 117; 1977) has been published in paperback by Cambridge University Press (Cambridge, UK) at £2.95.

● *Perceptions and Representations: The Theoretical Bases of Brain Research and Psychology*, by Keith Oatley (for review, see *Nature*, 277, 71; 1979), has been published in paperback by Methuen (London) at £3.95.

Depositing thin films

Nucleation and Growth of Thin Films. By B. Lewis and J.C. Anderson. Pp.504. (Academic: New York, San Francisco and London, 1979.) £27.50; \$57.

STUDIES of condensation of vapour atoms released in vacuum from localised sources can be dated from Knudsen's report in 1909 that mercury atoms did not condense on glass unless it was at a critical temperature, below -140°C ; an observation reinforced by R.W. Wood who found a critical temperature effect for condensation of cadmium atoms on glass. Wood believed that the vapour atoms were reflected and Knudsen attributed this to a failure of the incident atoms to adjust to the surface temperature. Certainly the effects of reflection have been found with atomic and molecular beams and in thermal transpiration experiments, but the phenomenon witnessed by Knudsen and Wood arose from temporary sojourn of incident Hg and Cd atoms on foreign surfaces and their re-evaporation. Thus, incident atoms can thermally accommodate to a surface and still evaporate if their binding energy is weak.

It was Langmuir in 1917 who provided this latter explanation, showing by experiment and deduction that an adatom, unlike an elastically rebounding atom, would have an average surface lifetime before re-evaporating. As the lifetime rose or the impact rate increased so could the adatom density rise and migrating atoms meet to form stable clusters. Thus, onset of condensation depended on both the incident atom impact rate and the substrate temperature for a given atom/surface pair. When the surface binding forces for the adatom were high condensation ensued without these critical conditions and molecular beam experimenters some 50 years ago precoated surface detectors with silver to condense weak Cd-beams. The nucleating effect was turned to advantage by the Bosch Company in Germany who made Zn-coated paper for capacitors using an Ag-seed layer, a process which in contemporary times has been applied to fingerprint detection.

The emergence of the electron microscope made it possible to observe the cluster growth in condensates and observe the onset of condensation with a vapour source mounted in the instrument. This technical advance and the growing interest in thin solid films in industry brought into being a world based group of 'nucleation' workers. In the same period studies of the condensation of gases, cryo-pumping, have grown and these have contributed knowledge on physical condensation mechanisms.

The book under review effectively collects together the results of studies of physical condensation during the past two

decades. The authors are known for their researches on film growth and physical properties. The reference text they have produced takes the reader through the logical sequence of physical adsorption, adatom migration, cluster growth and decay, nucleation kinetics and experiments on nucleation and epitaxial growth. The book is basically concerned with the physical condensation of atoms emitted at moderate thermal energies from a separate source, that is, non-equilibrium (or vacuum) evaporation. Thus, the title could be misleading particularly as films are now deposited by many different processes involving nucleation.

Deposition by sputtering is only briefly mentioned and an example given of film growth by ion beam sputtering. The high emission energy of sputtered atoms is referred to but other sources of energy transfer, such as backscattered energetic neutrals, are not considered. (On page 3 there is some confusion in use of the phrase "fairly high vacuum pressures" to describe the operating region for sputtering. Also this is not 10^{-1} to 10 torr as stated but 10^{-3} to 10^{-1} torr). However, this is basically a text dealing with atom impact at low energies in vacuum and we shall probably have to wait many years before all the conditions contributing to film growth by sputtering are unravelled although much is already known.

Immobilised enzymes

Immobilized Enzymes: Research and Development. Edited by I. Chibata. Pp. 284. (Wiley: New York, London, Sydney and Toronto, 1979.) £24.50; \$46.

IN recent years there have been many books published on various aspects of immobilised enzymes. Nevertheless, this book is a welcome addition, coming as it does from Dr Chibata and his colleagues. Dr Chibata, as Director of the Research Laboratory of Applied Biochemistry at the Tanabe Seiyaku Co. Ltd, Japan, has contributed greatly to the understanding of immobilised enzymes and cells, and to the development of industrial processes using these novel catalysts. He is therefore internationally recognised as a leading expert on the topics covered in this book.

The book contains four chapters, of which the first is a short introduction to the subject. The second chapter is devoted to the preparation of immobilised enzymes and cells, with comprehensive lists of the techniques that have been used and the enzymes and cells that have been immobilised. It concludes with a useful section on immobilised coenzymes. The next chapter describes the effects of immobilisation on the properties of enzymes and cells, such as the pH and temperature dependencies of their

Why do workers in this field usually refer to Frenkel for the adatom/lifetime relationship but rarely give credit for the basic concept to Langmuir and his equally important equilibrium (hyperbolic) law for adsorption? Langmuir's work is not referenced in this text.

This is not an easy work to follow — the subject is complex and full of detail — and a table defining the many symbols used would save constant back searching. Several typographical errors and some of sense occur in the text (the index in at least one case did not agree with the text and is inadequate). The book will be of value to those entering this specialist field. However, a note of warning — nucleation not only depends on substrate material and impact conditions but also on surface state. Research in this field can require ultra-high vacuum apparatus with mass spectrometers to determine the mass distribution of the incident and desorbed vapour and the residual gas composition, and instruments to determine the substrate surface condition after *in situ* cleaning, for example, by ion etching. It is as costly as it is complex.

L. Holland

L. Holland is a Professor in the School of Engineering and Applied Sciences, at the University of Sussex, Brighton, UK.

activities and the operational and storage stabilities in various conditions. The last and longest chapter is concerned with the use of these catalysts in chemical and food processes, and their analytical and medical applications. Naturally, there is considerable emphasis on processes developed by the authors but it is surprising to find so little about the production of high fructose-glucose syrups, which is by far the largest industrial application of immobilised biocatalysts. The final part of this chapter covers affinity chromatography and seems rather out of place in this book.

In the preface it is stated that the aim was to produce a comprehensive text, and to a certain extent the authors have succeeded. Unfortunately, presumably because of the interests of the authors, the book concentrates much more on the biochemical aspects and the engineering content is minimal. For instance, the section on reactors is very short and there is no consideration of the problems that arise from diffusional limitation of the reaction rate. As a result, the book tends to be rather descriptive and lacks a quantitative approach. Despite these criticisms, this book is highly recommended as a good reference text.

M. D. Lilly

M. D. Lilly is Reader in the Department of Chemical and Biochemical Engineering at University College London, UK.